Evolutionary Simulation Models:
On their character, and application to problems concerning the evolution of natural signalling systems

Seth Gwydion Bullock

Submitted for the degree of DPhil

University of Sussex

November, 1997

Declaration

I hereby declare that this thesis has not been submitted, either in the same or different form, to this or any other university for a degree.

Signature:

Acknowledgements

Many people have contributed to this thesis. I will not be able to list them all, but some of the most important cannot go unmentioned. First, this thesis could not have been completed without the support and encouragement of many people. Foremost among this group is the late Dr. Jane Beattie. Professor Graham Loomes stepped in to fill Jane's shoes and, like her, provided invaluable advice, a sense of humour, and a usefully orthogonal perspective on my work, in addition to enormously appreciated financial support.

I am indebted to the School of Cognitive and Computing Sciences for waiving my tuition fees for the duration of my doctoral studies, to those who provided me with teaching work during these studies, and to Phil Husbands for arranging a period of employment at the outset of these studies. I must also acknowledge the contributions made by the Baldwin Bewdley Trust and the Bromyard Grammar School Foundation, charitable organisations which provided me with £150 toward books, and three installments of £250 each towards living expenses, respectively.

My family have offered support throughout my studies. Particular mention must go to my grandmother who, in addition to taking an interest in my financial well-being, occasionally also provided aid of a more sartorial kind.

Various colleagues, both within COGS and beyond, provided the perfect critical audience, and contributed to the ideas presented here to a great extent. Particular amongst these were Guillaume Barreau, Pete de Bourcier, Ezequiel di Paolo, Berkan Eskikaya, Joe Faith, Inman Harvey, Nick Jacobi, Phil Jones, Ronald Lemmen, Giles Mayley, Jason Noble, Richard Vaughan, Henrietta Wilson and Mike Wheeler.

Henrietta Wilson, in particular, convinced me that evolutionary thought was not as unproblematic as I had at first imagined, and endured many hours of discussion, especially during recurrant periods of crisis. Mike Wheeler and Peter de Bourcier inculcated in me an interest in biological simulation, whilst Jason Noble and Ezequiel di Paolo thrashed out some of the thornier methodological and theoretical problems with me.

The exceptional supervision provided by Dave Cliff and Phil Husbands ensured that my studies proceeded with only infrequent interruption. The continual, unqualified, and unconditional positive reinforcement offered by Dave (despite the odd initial knock-back), and Phil's preternatural social urbanity, mark them as doyens of the first water.

More generally, I would also like to acknowledge the stimulating and interdisciplinary environment which exists within COGS; an environment which directly results from the energy and enthusiasm provided by (amongst other groups) the Artificial Life Reading Group, the Cognitive Philosophy Reading Group, and the first EASy MSc cohort, who unwittingly provided me with a useful forum for testing ideas at an early stage of development.

Over the last seven years, I have been aided and abetted in extra-curricular duties by a formidable phalanx of tireless devotees — Anthony Johnston, Sarah Duncan, Giles Duffy, and Richard Vaughan, take a bow.

Preface

A few of the chapters within this thesis have suffered a more concerted review process than the others as a result of their having been published or presented elsewhere…

Chapter 4

Much of this chapter was previously published as a technical report (Bullock, 1995a). Many thanks to Inman Harvey, Phil Husbands and Richard Vaughan for discussion of the issues addressed herein, to Julia Hutchison for her numeracy, and to Dave Cliff and two anonymous reviewers for comments on a previous draft.

Chapter 6

Much of this chapter was previously published as a paper co-authored with Dave Cliff (Bullock & Cliff, 1997). However, the text of this chapter and the results presented here are my own. The qualitative findings have been replicated independently by Dave Cliff using his own code. I thank Anthony Arak, Magnus Enquist, and Daniel Osorio for discussions, and Marian Dawkins, Henrietta Wilson, and an anonymous reviewer for help with previous drafts of this chapter.

Chapter 7

Much of this chapter was previously published as (Bullock, 1997). However, the text of this chapter includes extended discussion, and previously unpublished results. Thanks to Guillaume Barreau, Ezequiel di Paolo, Tim Guilford, John Maynard Smith, Jason Noble, Danny Osorio, Marian Stamp Dawkins, Michael Wheeler, and Henrietta Wilson, for discussion concerning the issues raised here, and help with previous drafts of this chapter. Some of the figures within this chapter were produced with the aid of Dave Cliff's graph-plotting utility, *eegplot*.

Thanks to all these reviewers and discussants, and, finally, special thanks to those few who trawled through my thesis for typos.

Evolutionary Simulation Models:
On their character, and application to problems concerning the evolution of natural signalling systems

Seth Gwydion Bullock

Abstract

Evolutionary simulation modelling is presented as a methodology involving the application of modelling techniques developed within the artificial sciences to evolutionary problems. Although modelling work employing this methodology has a long and interesting history, it has remained, until recently, a relatively underdeveloped practice, lacking a unifying theoretical framework.

Within this thesis, evolutionary simulation modelling will be defined as the use of simulations, constructed under constraints imposed by evolutionary theories, to explore the adequacy of these theories, through the modelling of an adaptive system's ongoing evolution.

Evolutionary simulation models may be considered to lie within the field of artificial life, since its concerns include theories of life, evolution, dynamical systems, and the relationship between artificial and natural adaptive systems. Simultaneously, evolutionary simulation modelling should be regarded as distinct from, yet complementing, existing evolutionary modelling techniques within the biological sciences.

The ambit of evolutionary simulation modelling includes those systems towards which one is able to take the *evolutionary perspective*, i.e., systems comprising agents which change over time through the action of some adaptive process. This perspective is broad, allowing evolutionary simulation models to address linguistic models of glossogenetic change, anthropological models of cultural development, and models of economic learning, as well as models of biological evolution.

Once this methodology has been defined, it is applied to a group of problems current within theoretical biology, concerning the evolution of natural signalling systems.

The ubiquity of natural communication is a well attested phenomenon. However, recently the utility of such communication within a world populated by neo-Darwinian selfish individuals has been questioned. Theoretical models proposed to account for the existence of signalling within the animal kingdom are reviewed, and evolutionary simulation models are constructed in an attempt to assess these theories. Specifically, models of the evolution of complex symmetry, and models of the evolution of honesty, are addressed.

Submitted for the degree of DPhil

University of Sussex

November, 1997

Contents

Chapter 1

Introduction

In 1837, twenty-two years before the publication of Darwin's *On the Origin of Species*, and over a century before the advent of the first modern computer, a piece of speculative work was published as an uninvited *Ninth Bridgewater Treatise*. The previous eight works in the series had been sponsored by the will of Francis Henry Egerton, Earl of Bridgewater, and a member of the English clergy. The will's instructions were to make money available to commission and publish an encyclopedia of natural theology concerning "the Power, Wisdom, and Goodness of God, as manifested in the Creation" (Brock, 1966; Robson, 1990).

The ninth publication in this series is noteworthy in that, unlike its eight forerunners, it neither sought to draw attention to miraculous states of affairs deemed unlikely to have come about by chance, and thus thought to be the work of a divine hand (e.g., the length of the terrestrial day, which seems miraculously suited to the habits of man and other animals), nor did it seek to reconcile scientific findings with a literal reading of the Old Testament (e.g., disputing evidence which suggested an alarmingly ancient earth, accounting for the existence of dinosaur bones, or promoting evidence for the occurrence of the great flood, etc.). In contrast to these apologetic efforts, the author of the ninth Bridgewater treatise produced what is, to my knowledge, the first instance of an *evolutionary simulation model*.

Since this thesis will address the topic of evolutionary simulation modelling it seems appropriate to begin with a brief account of this model. This will allow an introduction of the two concerns with which this thesis will be occupied. The first of these two concerns involves the definition of evolutionary simulation modelling. Chapters 2 and 3 will address the relative position that evolutionary simulation modelling occupies, contrasting it with work within the field of artificial life and the evolutionary sciences in an attempt to provide a coherent theoretical framework within which to embed the modelling practice.

The second concern will be to provide prototypical examples of the use of evolutionary simulation modelling to test scientific theories. The evolution of natural signalling systems will be chosen as a research area amenable to such an approach, and will be reviewed in Chapter 5, whilst Chapters 6 and 7 will offer specific models designed to assess particular theories within this field. The intervening Chapter 4 will serve as an interlude of interest to those who wish to turn evolutionary simulation modelling to more practical purposes.

As can be seen from the synopsis presented above, both conceptual and empirical issues will be addressed within this thesis. In addition, throughout this thesis, an historical perspective will frequently be adopted in an attempt to provide a coherent picture of the relationship between the work presented here and the studies which preceded it.

1.1 The Ninth Bridgewater Treatise & The First Evolutionary Simulation Model

The author of the *Ninth Bridgewater Treatise* was Charles Babbage, the designer of the difference engine and analytical engine (the first automatic calculating devices, and thus precursors to the modern computer). Indeed, in 1837 he was one of perhaps a handful of scientists capable of carrying out research involving both evolution and computational modelling. Even so, Babbage's model takes the form of a thought experiment, since neither the difference engine nor the more complex analytical engine were fully constructed until after Babbage's death in 1871. Furthermore, the notion of evolution employed by Babbage is distinctly underdeveloped compared with either that presented by Darwin over two decades later, or the evolutionary notions existing at the time of Babbage's writing (e.g., Lamarck, 1809). Despite these two hindrances, the model stands as a usefully simple case study with which to approach the character of evolutionary simulation modelling.

Babbage's (1837) model (see also Babbage, 1864, Chapter XXIX "Miracles" for a rather whimsical account of the model's development) was situated within what was then a controversial debate. It addressed the geological dispute between catastrophists and uniformitarians.

Prima facie this debate was internal to geology, since it concerned the geological record's potential to show evidence of divine intervention (principally in the form of support for the Old Testament accounts of the Creation and the Deluge). Catastrophists argued for an interventionist interpretation of geological evidence, taking discontinuities in the record to be indicators of the occurrence of miracles (violations of laws of nature). In contrast, uniformitarians insisted that in order to carry out scientific enquiry, the entire geological record must be assumed to be the result of unchanging processes. Allowing a role for divine miracles, the uniformitarians claimed, would render competing explanations equally valid. No theory could be claimed to be more parsimonious or coherent than a competing theory which invoked *necessarily inexplicable* exogenous influences in its account of the phenomena at issue.

The ramifications of this debate for empirical science were plain. For as long as discontinuities could not be accounted for in uniformitarian terms, catastrophists could justify the postulation of unaccountable external influences, and scientists interested in accounting for phenomena empirically would find such influences unpalatable as they rendered scientific explanation impotent.

Although this dispute had already been dealt something of a death blow with Lyell's (1830) publication of his *Principles of Geology*, the publication of the Bridgewater treatises and works like them evidences its slow demise. Only subsequent to the *coup de grâce* provided by Darwin's work on evolution did natural theology texts finally cease to be published (Brock, 1966).

Babbage's response to the catastrophist position was to construct what can now be recognised as a simple evolutionary simulation model. He proposed that a suitably programmed difference engine could be made to output a series of numbers according to some law (e.g., the integers, in order, from 0 onwards), but then at some pre-defined point (e.g, 100000), to begin to output a series of numbers according to some different law (e.g., the integers, in order, from 200000 onwards). Although the *output* of such a difference engine (an analogue of the geological record) would feature a discontinuity (in our example the jump from 100000 to 200000), the *underlying process* responsible for this output had been constant (i.e., the general law, or program, that the machine was obeying had not changed). The discontinuity was the result of the naturally unfolding mechanical (and computational) process. No external tinkering analogous to the assumed intervention of a providential deity had taken place.

Babbage thus tried to show that what might appear to be discontinuities were not necessarily the result of meddling, but could be the natural result of unchanging processes. In doing this he cultivated the image of God as a programmer, engineer, or industrialist, capable of setting a process in motion that would accomplish His intentions without Him intervening repeatedly. In Victorian Britain, the notion of God as draughtsman of an ‘automatic’ universe, one which would run unassisted, without individual acts of creation, destruction, etc., proved attractive. This conception was subsequently reiterated by several other natural philosophers (e.g., Darwin, Lyell, and Chambers) who argued that it implied “a *grander* view of the Creator — One who operated by general laws” (Young, 1985, p.148).

For the purposes of this introduction, what is of interest here is not the theological implications of Babbage’s work, nor the effect it had on the catastrophist/uniformitarian debate, but the manner in which Babbage mobilised his computational resources to attack a theoretical position. Babbage’s computational system was a simple analogue of a natural system (the geology of the planet) implemented mechanically. Babbage did not seek to capture the complexity of real geology in his system. Indeed the analogy between the difference engine’s program and geological processes is a crude one. However, the formal resemblance between the computing machine and the geological process is sufficient to enable a point about the latter system’s dynamics to be made. Babbage’s computing machine is thus clearly being employed as a *model*.

For now I shall award it the status of an *evolutionary simulation model* since it is implemented as an unfolding computational process. The necessary tightening of this notion will be deferred until the next chapter, but, for the moment, a few brief observations will serve to give a flavour of what is intended by the phrase.

First, that Babbage’s model is an unfolding computational process sets it apart from work in which models of dynamic change are constructed as mathematical proofs and are thus not evolutionary simulations. For example, Thomas Malthus’ (1798) work on population dynamics, in which he demonstrated that population growth would outstrip that of agriculture, was constructed using paper and pencil.

However, it is not merely the computational nature of Babbage’s model which ensures its status as an evolutionary simulation model. His reliance on the ongoing dynamic behaviour of his computational model, rather than on any end result it might produce, distinguishes it from modelling work which, although involving computational processes, uses computers as tools for solving what would otherwise prove to be intractable mathematical problems. For example the use of computers to discover digits of pi, or to iteratively solve the differential equations which might comprise a model of population dynamics will not count as *evolutionary* simulation modelling since the computational processes involved are merely the means of reaching a particular solution.

In contrast, the substantive element of Babbage's model is the evolutionary aspect of the simulation (i.e., the manner in which it changes over time). For the thought experiment which Babbage considers, his suitably programmed difference engine will, in principle, run forever. Its calculation is not intended to produce some end product, but rather the ongoing calculation is itself the object of interest.

1.2 Evolutionary Simulation Models

In Chapter 2 of this thesis, the class of evolutionary simulation models, which includes Babbage's system as its oldest member, will be outlined. In order to define this class, attention will be paid to two issues. First, what characteristics must a piece of evolutionary simulation work have to possess for it to be a *model*, and secondly what characteristics must a model have to possess before it can be termed an *evolutionary simulation* model. Dealing with these questions will require some preliminary groundwork.

Before criteria for membership of the class of evolutionary simulation models can be ascertained, some attention to the general constitution of artificial life, and to the scientific status of models will be necessary. To this end, Chapter 2 will begin with a brief (and selective) sketch of the field of artificial life. Subsequently, the scientific status of artificial life research will be discussed, and a typology with which to classify artificial life work will be introduced. Within this typology, the class of artificial life work which will be of central importance for the remainder of the thesis, evolutionary simulation models, will be defined, and criteria governing the constraints on such models will be outlined.

At this point, although a workable definition of a certain style of modelling will have been achieved, the factors constraining the applicability of this modelling approach to viable subject matter remain undetermined. Although Babbage's system models an abstract geological process, the majority of extant modelling work in this vein addresses biological subject matter. Should evolutionary simulation modelling be limited to problems within biology? What are the prospects for evolutionary simulation models which address linguistic, economic, or psychological subject matter? Discussion of the scope of application of evolutionary simulation modelling techniques will be necessary before such concerns can be properly assessed.

1.3 The Evolutionary Perspective

Chapter 3 will introduce the notion of the *evolutionary perspective*; a perspective which can be taken towards any system which comprises agents which change over time through the action of some adaptive process. It will be noted that any system towards which one is able to take the evolutionary perspective is a system which may be validly modelled using evolutionary simulation modelling techniques.

The appeal of the evolutionary perspective is very strong. However, a brief discussion will demonstrate that the factors responsible for its appeal have in some instances been fallacious. Primary among these misconceptions is the, sometimes unacknowledged, notion of an inexorable progress inherent in the evolutionary process.

The 'arguments from Design' employed throughout natural theology (e.g., Paley, 1802; Dawkins, 1986), and presented by the authors of the Bridgewater Treatises as supporting the existence of God are founded on the observation that many aspects of the natural world seem to be wonderfully complex, robust, economical, attractive, and generally suited to the existence and habits of man and other living creatures. That this is so is taken to be an indication of the existence of some coherent, conscious, benevolent, designer.

Subsequent to Darwin's formulation of an evolutionary account of natural history and the resulting gradual demise of natural theology, the very same observations of a well-adapted natural order, coupled with a naive appreciation of the evolutionary process as involving some directional element, have lead to a different, yet equally teleological, lay interpretation of the evolved world. The features of natural forms considered attractive from an anthropocentric perspective (e.g., beauty, economy, complexity, robustness, etc.), which were once thought to demand an explanation in terms of a creation myth, are now accounted for by the 'invisible hand' of evolutionary selection. Over evolutionary time, natural selection, and other evolutionary forces are assumed to change organisms from simple, fragile protozoa, into majestic, complex primates, cunningly fashioned insects, etc.; from lonely, blind, dumb, automata, to social, vocal, intellectual triumphs of evolutionary progress.

An account of this pervasive misconception (which afflicts evolutionary scientists as well as lay persons and academics from non-evolutionary disciplines) will be presented as both a cautionary point concerning (mis)application of the evolutionary perspective, and a useful context from which to approach the practice of artificial evolutionary design.

1.4 Prospects for Artificial Co-Evolutionary Design

One field in which the evolutionary perspective is currently enjoying a fashionable notoriety could be termed *artificial evolutionary design*. Chapter 4 will address the tension between the resolutely directional nature of the manual design process and the resolutely directionless character of natural selection. If, as has been argued in the previous section, evolution has no overarching progressive tendency, how can practitioners of artificial evolutionary design rely on artificial evolution to produce satisfactory solutions to design problems in reasonable amounts of time? Indeed, why would engineers expect evolution to be a powerful search process? Unsurprisingly the seeds of the attraction in viewing evolution as a directed search process stem from the same fallacious perspective described in the previous chapter, and ultimately, as was mentioned above, from a germ-line which stretches back to the apologism of natural theology.

As an instance of this mis-application of the evolutionary perspective, the notion of artificial co-evolutionary design will be introduced, its attraction outlined, and its actual utility assessed. It will be concluded that although there is a role for artificial evolution in designing artefacts, that role is a much more limited one than is typically imagined.

Rather than considering artificial evolutionary design to be harnessing the creative power of *natural* selection (or sexual selection, or other forms of natural evolutionary process), it will be proposed that artificial evolutionary design be considered as a form of *artificial* selection. By this I mean to compare artificial evolution as it is practiced by design engineers to artificial selection as it is practiced within animal husbandry. This change of perspective affords a more realistic conception of the limitations inherent in artificial evolutionary design, and neatly avoids the problems of the teleological fallacy outlined above by placing the designer centrally with respect to the evolutionary process, and entrusting her with the burden of directing it, rather than leaving this to some propitious ‘invisible hand’.

1.5 The Evolution of Natural Signalling Systems

For the remainder of the thesis, attention will be turned toward a class of animal behaviour to which the evolutionary perspective may lawfully be taken. Signalling behaviour is abundant throughout the animal kingdom (see Hauser, 1996, for a recent review of the biological signalling literature). The diversity of communication systems is huge, and styles of signalling vary along many dimensions. Some signalling systems involve relatively conservative signals, e.g., the intracellular signals which regulate cell behaviour, or the signals of playful intent used by many juvenile animals to ensure that their actions are interpreted as play behaviour rather than attempts at attack or copulation. In contrast, some signalling systems involve relatively extravagant signals, e.g., signals of aggressive intent, mating displays, signals of unpalatability, etc. In addition to these uncontentious examples, some signalling systems seem less straightforwardly communicative, e.g., camouflage and crypsis. Finally, the unique phenomenon of human language, so central to our everyday notions of communication, seems very far removed from any of the comparatively crude semaphores with which the rest of the animal kingdom make do.

Biologists faced with this profusion of differing systems evidently have their work cut out accounting for the particular character of each. However, until recently workers in this area were at least assured of the adaptive value of such signalling behaviours and thus had no cause to question the evolutionary function of communication. Communication between members of a group of animals with some confluence of interest was considered beneficial in as much as it beneficially co-ordinated the behaviour of the group. Such co-ordination was reasoned to confer advantages, in terms of efficient resource allocation, which groups of non-signalling, less well co-ordinated animals would not enjoy. Hence signalling groups would proliferate at the expense of their taciturn competitors.

Such group-selectionist accounts of co-operative behaviour fell into disrepute during the mid-sixties as the result of the observation that what was beneficial for an individual was not necessarily beneficial for the group it belonged to (Williams, 1966; Hamilton, 1964). Adaptive explanations at the level of the individual or gene were proposed as more powerful and parsimonious than those at the level of the group, and group selectionist thought was characterised as “sloppy” (Sober, 1994, p.xii). These were the beginnings of a more individualistic perspective on evolutionary change (as popularized in Dawkins’ *The Selfish Gene*) which would not tolerate accounts of communication as adaptive merely at the group level.

If signallers act selfishly, it was mooted, why should they signal anything other than manipulative lies, and conversely, if observers act selfishly, why should they bother to pay attention to such unreliable signals. However, although it was feared that the selfish actions of individuals might compromise the stability of natural signalling systems, such systems, as noted above, appeared to be the frequent products of evolution. Evolutionary theorists were thus left with an explanatory vacuum to fill. It is the theories which stepped into this breach which form the subject matter of the empirical portion of this thesis.

If signalling systems exist to promote the needs of the individual (or the needs of individual genes), one might expect the signals involved to be exploitative as evolutionary success can only be achieved at the expense of others. In a pair of seminal papers, Dawkins and Krebs (1978; Krebs and Dawkins, 1984) proposed that signallers and the observers of signals might experience a conflict of interest in that they may want different things from the signals they exchange. Signallers benefit from signals which manipulate their receiver in some fashion advantageous to the signaller. In contrast, receivers of signals benefit from a source of relevant, veridical, information which would otherwise be unavailable. The authors propose that this conflict results in an evolutionary struggle, or arms race (Dawkins & Krebs, 1979).

From this perspective, although the signalling systems we observe throughout the natural world seem intricately balanced, signallers wholeheartedly calling, warning, informing, etc., whilst receivers attend, and respond, this stability is illusory. Over evolutionary time periods the one-upmanship predicted by Krebs and Dawkins would become apparent.

For example, propositioned mates might discover, and pay attention to, some indicator which predicts the characteristics of a potential mate of interest to them (perhaps the colour of a male bird's ruff is a good indicator of the male's virility). The indicator (particular coloured ruff) might become a compelling signal which reliably elicits mating opportunities. As a result mating displays which feature this indicator will gradually proliferate in the male population. But this proliferation might occur at the expense of the indicator's predictive power (the coloured ruff is now sported by all manner of males including, perhaps, impotent suitors). The ubiquity of such a signal reduces its informative content (the presence of a particular coloured ruff ceases to be a strong indicator of virility), forcing the choosy mates to abandon it in favour of some further method of eliciting the information they require (perhaps the size of the ruff is better correlated with virility), and the cycle starts again.

Such signals have been termed conventional (Dawkins & Guilford, 1991) in that their utility relies on the maintenance of a convention. Such signals bear information only arbitrarily, and are sensitive to the effects of signallers that break with the convention, and thus degrade their informational content.

Conventional signalling models propose that dishonest individuals will remain at a low frequency in the signalling population, parasitic upon the honesty of others. If the conventional relationship between the signals and relevant states of affairs is compromised to too great an extent the signalling system may collapse. However, negative feedback within the signalling system acts to prevent this through increasingly degrading the signal's worth as it is increasingly abused, and thus both discourages the increasing proliferation of dishonest strategists, and encourages signallers to establish alternative conventional signalling systems.

For example, a conventional account of the evolution of aposematism (emetic unpalatability) and mimicry (merely resembling an unpalatable prey type without investing in unpalatability itself) might claim that a strong conventional association between a particular appearance (e.g., red spots on black wings) and unpalatability would have to exist prior to the success of deceitful mimetic adaptations. Further, as the strength of the convention upon which the success of the mimic depends is itself dependent on the degree to which the convention is observed, the success of the mimetic strategy will decrease as mimics proliferate within the population. Thus conventional signalling models predict a low level of cheating, maintained by an evolutionary negative feedback.

The implications and further details of this position will be addressed in Chapter 5. For the time being it will suffice to present this conventional account as establishing a foundation upon which two further developments have been built. The first involves the role of sensory and motor biases in signalling systems (e.g., Ryan, 1990).

1.6 Biases and Signal Evolution

Since animals typically utilise their sensory and motor equipment for all manner of activities, it has been proposed that certain biases or predispositions may exist within those systems prior to their becoming used to effect or detect signals. For example, the visual system of an animal which uses sight to find prospective mates, might also be used to forage for food. In order that the animal forages effectively, some adaptation of the visual system to this task might have occurred over evolutionary time. Such an adaptation may be ripe for exploitation. For example, a tendency of females to orient toward stimuli of a certain colour, which has evolved as a foraging adaptation, might be exploited by suitors during mating displays.

This situation is similar, but subtly different from the conventional situation described above in that, unlike the response strategy within a conventional signalling system, the signal recipient's response is not as free to adapt under selective pressure. Although exploitation of the observer's sensory bias by signallers might promote the exaptation of the bias (i.e., signaller behaviour might exert a selective pressure, which, in a conventional signalling system, would tend to result in a change of response behaviour on the part of the signal receiver), the role of the receiver's sensory bias as a useful adaptation in an alternative context may ensure its persistence in the population. Consider again the previous example in which a sensory bias which causes females to attend to areas of the visual field which are a certain colour is exploited by male suitors in order to gain female attention. Although this exploitation might result, on average, in females mating with lower quality males, the sensory bias may persist due to its utility in prey detection.

Furthermore, as this example illustrates, for signalling systems founded on sensory biases, the relationship between signal and informational content is not arbitrary, but is constrained by the nature of the pre-existing bias within the system. Whereas conventional signals are considered to carry information arbitrarily (e.g., that red means stop, and green means go, is not taken from scripture, but is settled upon through some more capricious process), the sensory bias described in the above example could not be exploited by signallers using an arbitrarily coloured display. Rather the form and function of the signal are intimately related to that of the receiver's dispositions.

Chapter 6 will present an evolutionary simulation model which assesses a hypothesis which proposes the evolution of symmetry as an exploitation of receiver bias (Enquist & Arak, 1994). The authors of the hypothesis claim that the complex symmetry of many displays in the natural world are the result of evolutionary pressures to exploit a bias inherent in all visual systems which favours signals which are invariant under many transformations. Such signals, it is claimed, are easier to discriminate from irrelevant stimuli. I present an evolutionary simulation model which examines how this hypothetical bias might operate, and the subsequent findings, which suggest that rather than promoting complex symmetry, in reality the bias promotes simple homogeneity.

This chapter will also assess some of the methodological issues involved in evolutionary simulation modelling in an attempt to discover why Enquist and Arak (1994) failed to discover the true nature of the bias they were exploring.

1.7 Honesty and Signal Evolution

Until now, the post-group-selectionist account of signal evolution which I have described has involved the potential for cheating. From the perspectives outlined above, signalling systems are considered to be inherently prone to dishonesty. However, one theory which has been under development for the past 20 years, and has attracted a great deal of attention from evolutionary theorists, proposes that many, if not all, signalling systems are inherently honest (Zahavi, 1975, 1977a, 1977b, 1987, 1993).

Zahavi suggests that the cost that signallers inevitably pay in producing a signal stabilises its honesty. This notion, which has been dubbed the ‘handicap principle’, is subtle, and attempts to formalise it have occupied many evolutionary theorists over the last few decades (e.g., Enquist, 1985; Grafen, 1990a, 1990b; Maynard Smith, 1976, 1991). A detailed exposition of its many facets must be postponed until Chapter 5. However, a brief sketch of the main thrust of Zahavi’s argument will be attempted here.

Those that stand to gain most from bluffing, cheating, treachery, etc., are those that cannot afford to make good their claim. They fear being put to the test. They claim to be healthy, wealthy, hungry, strong, agile, knowledgeable, trustworthy, etc., but do not wish to demonstrate these qualities prior to gaining copulation, food, security, admission, reprieve, or whatever resource it is that they are begging for. Since the production of an arbitrary deceptive signal is cheaper than obtaining the quality it dishonestly advertises, such individuals will be tempted to cheat, bluff, or lie. As noted above, conventional models of such signalling predict just such cheating, maintained at a low level by an evolutionary negative feedback mechanism.

In contrast Zahavi’s contention is that the very act of signalling is a test to which dishonest individuals are being put. Not only this, but it is a test that the dishonest signaller must surely fail.

This contention is underwritten by the following observation. Those that *are* healthy, wealthy, etc., can more easily afford to make signals to that effect than those which are not. Since this is the case, over evolutionary periods of time, the forces of signal selection (Zahavi, 1987) will ensure that signals come to vary in proportion to the quality that they advertise, and that a signal’s form will not be arbitrarily related to the information it carries, but will be significantly influenced by it.

An example may help to clarify this notion. In scenarios in which one individual’s wealth is a deciding factor in another’s decision making (as some would have it be in human mate selection, for example), it pays the decision maker to exploit any accurate indicator of wealth. According to the handicap principle, such an indicator would be the extravagance of the individual’s spending. Since only the wealthy can afford to waste money, a wanton disregard for expenditure can be taken as an indicator of wealth (Zahavi, 1987). Signals demonstrate the ability of an individual to bear a handicap. As such they are honest indicators of the degree to which the signaller possesses the resource which it is willing to waste through signalling.

Zahavi's position is that the honesty of many signalling systems typically considered to be of a conventional nature is in fact maintained by some form of handicap effect. He lists intracellular signals, ritualised movements, vocal signals, colour patterns, mimicry, and altruistic acts as honest signals stabilised by their cost (Zahavi, 1977b, 1987), even going so far as to suggest that the widening of the eyes which he claims occurs when humans gaze at a loved one is an honest indicator of affection maintained by the handicap of decreased acuity during such bouts of wide-eyedness (Zahavi, 1987, p.317). Thus Zahavi clearly feels that the handicap principle is of near universal applicability. This position appears to motivate his enthusiastic suggestion that "as a method, the assumption that a signal is involved with cheating usually leads to a dead end while the search for the honest meaning may lead one to exciting new discoveries" (Zahavi, 1987, p.317).

In Chapter 7 a series of evolutionary simulation models will be presented in an attempt to clarify the still problematic notion of the handicap principle. The points of departure for these models will be the work of Grafen (1990b) and Hurd (1995), both of whom attempt to demonstrate the integrity of the handicap principle analytically.

First Hurd's (1995) discrete action-response model will be extended and subsequently rendered as an evolutionary simulation model in order to show that the benefits accrued by signallers from their signalling behaviour are as important as the costs incurred. This model will also serve to demonstrate the compatibility of conventional signalling accounts with those couched in terms of the handicap principle.

Subsequently an analogous extension to Grafen's (1990b) continuous model of the handicap principle in the context of sexual selection will be undertaken. This model will demonstrate the extent to which the findings of the discrete model can be generalised to models involving continuous variables. It will be shown that the implications of these models demand a recasting of Zahavi's handicap principle in cost-benefit terms. This recasting will go some way towards reconciling the differences between conventional and handicap theories and will also enable the conditions favourable to each to be determined.

1.8 Summary

This thesis will establish a framework within which the dynamic behaviour of natural systems, comprised of agents subject to some adaptive process, can be modelled as ongoing activity. Such modelling will be termed evolutionary simulation modelling since the simulation of a system's evolution will be central to any such modelling.

Thus described such modelling can be contrasted with conventional models of adaptive behaviour, e.g., game theoretic models within economics (von Neumann & Morgenstern, 1944), and biology (Maynard Smith, 1982), in which the transient behaviour of models prior to the achievement of some stable state, if it is considered at all, is typically considered to be secondary to the nature of the possible stable states admitted by the model.

In addition, such evolutionary simulation modelling may profitably be compared with current movements towards dynamic, or evolutionary, perspectives within many traditionally non-evolutionary disciplines, e.g., evolutionary linguistics (Pinker, 1994), evolutionary economics (Dosi & Nelson, 1994), and developmental psychology (Thelen & Smith, 1994).

Subsequent to the establishment of a framework capable of supporting evolutionary simulation modelling, the validity of theories concerning the evolution of natural signalling systems will be addressed through the construction of such models. Specifically, models will be constructed in order to test hypotheses concerning the role of biases within signalling systems, and problems of honesty within such systems.

The role of evolutionary simulation modelling, as outlined within this thesis, is to enable the assessment of a wide range of theories which purport to account for the behaviour of various natural systems. Since it is the case that such systems are studied to a large extent independently by specialised groups of academics, each with its own methodology, philosophy and ideology, evolutionary simulation modelling, like other general purpose modelling techniques, is by its nature inherently interdisciplinary. In addition to performing the standard duties of a concluding passage, Chapter 8 will close with some thoughts on the nature of interdisciplinary work.

Charles Babbage's (1837) unsolicited attempt to alleviate certain problems within geology, the work with which this introduction opened, in many ways exemplifies the precarious, interdisciplinary role of evolutionary simulation modelling. His use of tools which were perceived to be ill-suited to the job (Brock, 1966) earned his model little credence (Young, 1985). Progress in this respect seems imperative if evolutionary simulation modelling is to develop successfully.

Chapter 2

Evolutionary Simulation Models

The title of this chapter is not intended to be a neologism. Neither 'evolutionary simulation modelling', nor 'evolutionary simulation model(s)', will be capitalised or abbreviated within this thesis since each is intended to be a descriptive phrase, referring straightforwardly to the use of computer simulations as evolutionary models of adaptive systems. However, these phrases do have a neologistic flavour, since they are necessarily shorthands. By evolutionary simulation modelling, I do not wish to refer to just *any* use of computer modelling within evolutionary biology, nor *any* evolutionary simulation design within artificial life. However, subject to certain provisos to be detailed within this chapter, certain pieces of work within either of these two practices might be legitimately classified as evolutionary simulation modelling.

As a crude first approximation, the phrase can be thought of as capturing the notion that some simulations may be used as models of adaptive processes, and, furthermore, that it is the *evolution* of those simulations (i.e., the manner in which those simulations change over time) which is their substantive element. It is the *dynamic behaviour* of such simulations which may be used to dispute or uphold the validity of evolutionary theories.

Such a characterisation of evolutionary simulation modelling clearly accommodates a style of computer simulation which could be employed by theoretical biologists. Simultaneously, the same characterisation of evolutionary simulation modelling just as clearly accommodates the use of artificial life techniques as evolutionary *modelling* tools. However, as this chapter unfolds, much of computer simulation within biology will be found to be of a different style to that motivated under the nascent auspices of evolutionary simulation modelling, whilst much of the evolutionary simulation carried out within artificial life will be found to fail to achieve the status of scientific modelling.

In an effort to clearly specify what is meant by these statements, the body of this chapter stands as a longhand exposition of its title, seeking to define a role for evolutionary simulation modelling which accounts for the influence of both artificial life and evolutionary biology.

A Naive Assumption

Mutually supporting arguments are often difficult to construct, each one requiring the others to be firmly in place before it can be securely presented. A typical solution to such a dilemma is to initially provide some scaffolding with which to support the whole construction. Once the thesis has been presented in its entirety, it simply remains to demonstrate that it can survive the gradual removal of the supporting material.

In contrast, the arguments presented within this chapter threaten to be mutually *defeating*, since the character of evolutionary modelling within biology is often presented in *opposition* to that of evolutionary simulation within artificial life. Before it is possible to show that these two practices have, as Bonabeau and Theraulaz (1996) put it, a “non-empty intersection”, and furthermore that this intersection defines a scientific tool with which problems within many diverse evolutionary fields may be addressed, some equivalent scaffolding must be in place, in order to separate the two antagonistic arguments.

This separating scaffolding takes the form of one naive assumption, namely, that evolutionary simulation modelling does not take place, and has never taken place, within biology, but is properly considered to be a practice lying exclusively within the field of artificial life. This assumption is patently incorrect, and potentially incoherent, since the overlap between artificial life and biology is considerable. Although computer modelling within biology has typically been of a style different to that motivated within this thesis, there have no doubt been instances of what could be classified as evolutionary simulation modelling within biology. Some of this work predates artificial life (e.g., Babbage, 1837), whilst some has been more recent (e.g., Krakauer & Johnstone, 1995).

However, for the time being the assumption that evolutionary simulation modelling is specific to artificial life will allow a clear account of its makeup to be constructed without requiring the constant qualification of statements, or the development of elaborate inter-related accounts comparing and contrasting approaches within biology with those within artificial life.

2.1 Chapter Outline

Evolutionary simulation modelling clearly stands in some relationship to both *evolutionary modelling* within biology, and *evolutionary simulation* as it is practiced within artificial life. This chapter will clarify the nature of these two relationships. Since the former relationship is most straightforward (primarily as a result of the relative maturity of evolutionary modelling within biology), most of this chapter will deal with the relationship between evolutionary simulation modelling and artificial life. The argument presented within this chapter is that whilst evolutionary simulation modelling has benefited enormously from the advent of artificial life modelling *techniques*, it has been hamstrung by artificial life rhetoric, dogma, and wide-eyed over-ambition.

Since the development of new modelling techniques within artificial life has often been associated with the development of new perspectives on what is being modelled (e.g., Kauffman, 1993), arguments in favour of the adoption of such modelling techniques have often been accompanied by, and confused with, arguments motivating the adoption of new philosophical positions, or conceptual schemes (e.g., Ray, 1994). Since evolutionary simulation modelling, although bearing strong resemblance to the practices carried out within several research programmes within artificial life, shares few of the ideological tenets which are espoused by their authors, particular care will be taken to outline the differences between evolutionary simulation modelling and other research programmes within artificial life.

This will be achieved though a protracted process of elimination. The repeated application of increasingly fine-grained methods of systematically distinguishing between different classes of artificial life practice will be used to identify the necessary and sufficient conditions for a piece of research to count as evolutionary simulation modelling.

This process of elimination will proceed as follows. First the strong/weak dichotomy sometimes used to distinguish between research agendas within artificial life (and artificial intelligence) will be dismissed as too blunt an instrument for the purposes of this chapter. Subsequently, a three-way distinction between aspects of artificial life which may be roughly characterised as philosophical, scientific, and engineering, will be presented. This will enable evolutionary simulation modelling, and other artificial life modelling work, to be distinguished from, on the one hand, the exploration of conceptual schemes within artificial life, and, on the other, the exploration of new design techniques within artificial life. To the extent that disciplines other than artificial life involve these three aspects, they too may be arranged within a similar hierarchy.

Amongst those researchers interested in employing artificial life techniques as scientific tools, one can identify an axis which promotes the use of *evolutionary simulations* as models with which to examine evolutionary phenomena. The multiple arguments in favour of this practice, although similar, may be discriminated between on the basis of the role evolutionary simulations are proposed to play with respect to (i) the natural sciences concerned with the relevant evolutionary phenomena, and relatedly (ii) the existing techniques with which they are modelled.

The various positions of these proponents can crudely be characterised as ranging from (a) the advocation of a *conservative* use of simulation models as an (often-claimed inferior) alternative to existing evolutionary modelling techniques, to (b) the advocation of a *radical* use of simulation models to explore aspects of evolutionary processes which are claimed to be necessarily opaque to existing modelling techniques.

Evolutionary simulation modelling, as it is presented within this thesis, is neutral with respect to this conservative/radical axis. However, whilst the conservative use of evolutionary simulation modelling is relatively uncontentious, proponents at the radical pole of this continuum risk the same conflation of modelling technique with philosophical position which dogs much of would-be artificial life modelling. For example, the role of simulations which involve notions of emergence, self-organisation, complexity, etc., will only be resolved as and when the theoretical utility of these notions becomes clearer.

By comparison, clear methodological differences between evolutionary simulation models and closely related evolutionary modelling techniques inspired by the application of game theory to evolutionary phenomena (Maynard Smith, 1982) have equally clear implications for modelling practices employing each technique. In order to assess the differences in perspective which each affords, the naive assumption which has been in place throughout the chapter will be relaxed, and simulation modelling as it is practiced within the biology community will be discussed.

2.2 Artificial Life

Between the publication of Babbage's (1837) evolutionary simulation model of geological discontinuities, with which the previous chapter opened, and the emergence of artificial life in the late nineteen-eighties, very few pieces of what could, retrospectively, be classed as evolutionary simulation modelling were published.

For the greater part of this period computing machinery was hardly more advanced than the machines designed by Babbage in the first half of the 19th century, but with the post-world-war development of the first modern electronic digital computers came a surge of research into computational theory, some of which addressed biological phenomena.

Seminal work by mathematicians such as Turing (1952) on reaction diffusion models of morphogenesis, von Neumann and Burks (1966) on automata theory and self-replication, McCulloch and Pitts (1943) on the logic of neural circuitry, Holland (1975) on the formal properties of adaptation, and Walter (1963) and Ashby (1956) on cybernetics, involved the application of logic, mathematics, robotics, and control theory to essentially biological problems.

The above-cited pieces of research (with the exception of Babbage) are now recognised as the intellectual precursors to the field which has come to be known as artificial life. Although, more proximally, artificial life can be considered to be the offspring of artificial intelligence (see Brooks, 1991, and Steels, 1994, for accounts of artificial life's relationship to artificial intelligence), it is becoming increasingly apparent that the work published under the artificial life rubric (e.g., models of morphogenesis, cellular automata models, behaviour based robotics, the simulation of adaptive behaviour, etc.) has inherited much of its method, and some would say madness, either directly, or circuitously, from these mid-century pioneers.

Of these original and inspirational pieces of research, techniques for implementing adaptive processes within computers have had the most direct impact on the development of evolutionary simulation modelling. Most important of these techniques are computer algorithms which implement some analogue of evolutionary adaptation.

Here, genetic algorithms (Holland, 1975) will be taken as prototypical of such adaptive algorithms, but several alternatives have been developed (evolutionary strategies, Rechenberg, 1973; genetic programming, Koza, 1993), and analogous work implementing ontogenetic, rather than phylogenetic, adaptation (i.e., learning) have also had a parallel influence within modelling paradigms concerned with development, rather than evolution (Rumelhart & McClelland, 1986; McClelland & Rumelhart, 1986).

Genetic algorithms (GAs) were developed by Holland during the 60s, and 70s, as an abstract model of evolutionary adaptation in natural systems. Their use was initially limited, but has grown rapidly during the last decade or so, resulting in the recent emergence of a wide variety of work within the genetic algorithm community (see chapter 4 for more details, and Mitchell, 1996, for a recent review of the subject).

A healthy proportion of genetic algorithm work within artificial life involves harnessing this artificial evolutionary process in order to automatically design artefacts (e.g., Hillis, 1990, applied GAs to the task of designing efficient sorting algorithms; Husbands, 1993, applied GAs to the problem of producing efficient schedules for production plants; Harvey, Husbands, & Cliff, 1994, applied GAs to the design of control architectures for mobile robots; Thompson, 1996, applied GAs to the task of evolving reconfigurable electronic circuits to perform a variety of tasks). However, its role within evolutionary simulation modelling is more in line with that for which it was initially intended, i.e., as a model of natural evolutionary adaptation.

Before an attempt is made to locate evolutionary simulation modelling within artificial life, a prototypical example of the application of this style of modelling to a biological problem will be presented.

In 1987, Hinton and Nowlan employed a genetic algorithm in an exploration of the Baldwin effect. This effect was hypothesized by Baldwin (1896) to increase the efficiency of natural selection through the effect of individual learning by organisms, but did not require the inheritance of acquired characteristics for this increase in efficiency to take place.

Hinton and Nowlan's (1987) study of the Baldwin effect took the form of a computer simulation which implemented a period of artificial evolution in which a population of artificial agents underwent reproduction with variation and heredity, and were subsequently selected for on the basis of an artificial criterion of fitness. In many respects this artificial evolutionary paradigm was similar to the *in vitro* evolutionary experiments carried out in many evolutionary biology laboratories. However, Hinton and Nowlan were able to explore the effects of varying the conditions under which their agents evolved in a manner which would prove very hard to implement in the laboratory.

The authors attempted to demonstrate that the strength and character of the Baldwin effect could be ascertained through the implementation and exploration of a genetic algorithm simulation.

The scenario considered by Hinton and Nowlan was, by their own admission, both extreme and simple. An artificial organism which possessed a phenotype consisting of 20 traits was envisaged. The absence or presence of each trait was coded for by one of 20 genes. One particular combination of the presence and absence of these 20 traits was specified to confer a reproductive advantage upon such an organism. The remaining combinations of present and absent traits were all specified to confer an equal but lower rate of reproduction. An evolutionary process faced with such a scenario can do no better, on average, than random search. However, evolution coupled with an organism's learning ability should discover the fitter phenotype more efficiently, if Baldwin's proposed evolutionary factor is effective.

The algorithm designed to test this hypothesis consisted of a number of elements. First, a population of artificial organisms was constructed. The genotype of each artificial organism consisted of a string of twenty characters (genes), each of which was drawn from an alphabet of three symbols (alleles) denoted, 1, 0, and ?. A particular allele at a particular location (loci) on the genome indicated the innate presence (1), innate absence (0), or unspecified status (?) of the trait coded for at that loci. Innately specified traits were static throughout an organism's lifetime, whereas traits with an unspecified status were subject to a learning algorithm during an organism's lifetime.

Learning was a simple affair. Each organism was allowed a maximum of 1000 attempts to learn the beneficial combination of traits. Each attempt consisted of each unspecified trait being randomly determined as either present or absent. If this new assignment resulted in an organism with the combination of present and absent traits which had been specified to confer upon it a reproductive advantage, the organism was awarded one point for every remaining attempt (i.e., if an organism achieved the beneficial combination of traits at the 400th attempt it would score 599 points).

A population of organisms was reproduced by choosing pairs of parental organisms and mixing their genotypes in order to create an offspring genotype which shared some characteristics with each of its parents. Parent organisms were chosen at random with a bias towards those with the highest scores, i.e., those which had spent most of their lifetime with the correct specification of traits. Those organisms who had never possessed this advantageous combination were the least likely to be chosen. Conversely, those organisms possessing this combination of traits from birth were the most likely to be chosen.

Thus, the fitness of an organism was dependent on two factors. First, a genotype with an ‘incorrectly’ specified innate trait would never enjoy an increase in fitness. Secondly, organisms with no incorrectly specified traits, would benefit from learning to the extent that they had fewer unspecified traits. These two factors can be thought of as ensuring that learning superimposes a gradient on a fitness landscape which otherwise would have been flat save for a sharp peak representing the combination of traits which enjoyed increased fitness. This gradient allowed evolution to gradually, but unproblematically, proceed from any genotype featuring no incorrectly specified innate traits to the genotype featuring the correct combination of genetically specified innate traits.

Hinton and Nowlan’s simulation did indeed result in convergence upon the optimal genotype at a rate much faster than could be achieved without learning (although a low level of unspecified genes remained in the population, see Harvey, 1993, for an explanation of this phenomenon in terms of genetic drift). Further work in the same vein as Hinton and Nowlan (1987) has gone some way towards identifying aspects of evolutionary landscapes which impact on the efficiency of the Baldwin effect. Issues such as the correlation between genotype space and phenotype space (i.e., the degree to which something which is easily learned is also something which can easily be genetically assimilated), the role of detrimental fitness effects due to learning, and the effect of differences in the character of the fitness landscape prior to the effects of learning, are all being explored (see, e.g., Mayley, 1996, and additional papers in the special issue of the *Evolutionary Computation* journal on Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect, edited by Turney, Whitley & Anderson).

As mentioned above, throughout the following discussion of the role of evolutionary simulation modelling within artificial life, Hinton and Nowlan’s (1987) study of the Baldwin effect should be kept in mind as a prototypical piece of such research.

2.3 Strong vs. Weak Artificial Life

A distinction between strong and weak artificial life has been made by Sober (1996) and Pattee (1996) in an analogous move to the distinction made by Searle (1980) between strong and weak artificial intelligence. For these authors, the terms strong and weak distinguish between research concerning artificial systems which are claimed to *possess* a quality (intelligence or life), and artificial systems which claim to increase our understanding of a quality (intelligence or life) through some systematic *similarity* with the natural systems which possess it.

2.3.1 Artificial Intelligence

The distinction between strong and weak artificial intelligence arose as a result of the adoption of a new analogy for the mind — the computational metaphor. A mind, it was claimed, could be considered as a complex machine, receiving inputs, processing those inputs (perhaps storing them or retrieving stored information during this processing), before outputting the information in the form of action.

The power of this metaphor was very strong. Whereas previous models of mind (e.g., Plato's comparison of the mind to a wax tablet), had been useful in *explicating* mental phenomena, it had never been suggested that they could *exhibit* mental phenomena. With the formulation of the Physical Symbol System Hypothesis (Newell & Simon, 1976) came the possibility that, rather than resembling the mind in some abstract manner, a computer could actually *be* a mind.

This hypothesis characterised cognition as symbol processing. What minds were doing could be viewed as structure-sensitive computational processes manipulating complex symbolic entities constructed systematically from primitive symbolic entities. By analogy with natural languages and computational languages, cognition could be said to involve a “language of thought” (Fodor, 1975).

2.3.2 Artificial Life

The analogous move made within artificial life is less clearly presented in the literature, but takes the form of various functionalist claims about the nature of life (e.g., Bedau, 1996). If life, like mind, can be claimed to be an abstract property of certain organisations, then life can be realised within media other than carbon-based chemistry. However, a list of consensually agreed-upon criteria with which to distinguish the class of living organisations from the class of non-living organisations has yet to be compiled. In addition, debate concerning candidate criteria has a long history and is still controversial (Boden, 1996b).

Partly as a result of this lack of consensus regarding the concept of life, much work within artificial life is hard to place with respect to the strong vs. weak dichotomy. Some research concerns systems which are neither claimed to be alive, nor claimed to offer any advance over previous notions of what it is to be alive. For example, evolutionary roboticists strive to produce robots which are robust solutions to difficult design constraints through the harnessing of some artificial evolutionary design process (e.g., Cliff, Harvey, & Husbands, 1993). Such work is a central aspect of artificial life, yet is hard to square with the strong/weak distinction. In contrast, some artificial life work seems to fall across the strong/weak divide. For example, “synthetic biologists” (Ray, 1994) and “synthetic ethologists” (MacLennan, 1991) both realise, *and* study, artificial systems which are claimed to potentially be alive (contingent upon their satisfying “formal criteria for life” MacLennan, 1991, p.638). Such researchers insist that the ‘inoculation’ of life into the medium of a digital computer (strong artificial life) is itself a technique with which to elucidate the nature of life (weak artificial life), i.e., for such researchers strong and weak artificial life merge into one pursuit.

Thus, whilst the strong/weak distinction may be useful as a conceptual grindstone on which to sharpen philosophical theories of life (as its progenitor has been useful within cognitive philosophy), it offers little methodological insight with which to guide the development of artificial life research. Since artificial life appears to be replete with philosophical cant, yet bereft of methodological rigour (Elton, 1994; Miller, 1995), an alternative classification system will be presented within which evolutionary simulation modelling can be located with respect to other scientific, philosophical, and engineering practices within artificial life.

2.4 A Class System for Artificial Life

The tripartite class system presented below is a rough hierarchy. It is not intended to provide hard and fast categories. Research programmes, and in some cases individual pieces of artificial life research, will sometimes reach outside these three classes, and will frequently fall across the boundaries between classes. However, this system will be structured enough to provide a clear home for evolutionary simulation modelling.

Each class will be described and prototypical research falling within its remit will be detailed. Subsequently, constraints which govern the work within each class will be presented.

2.4.1 High Class Artificial Life

Research within this class of artificial life concerns itself with the development and exploration of conceptual schemes. As such it is most naturally understood as philosophy. Indeed, much of the work which exists within this class is published under a 'philosophy of artificial life' rubric (e.g., Boden, 1996b).

Just as artificial intelligence supported computational accounts of cognitive philosophy (e.g., Churchland, 1984), High Class artificial life impacts on alternative schools of philosophy. Non-Cartesian philosophical positions which advocate the centrality of notions such as embodiment and situatedness, originally due to continental philosophers such as Heidegger (1926), have proved amenable to treatment within High Class artificial life (e.g., Wheeler, 1995; Hendriks-Jansen, 1996), as have reinterpretations of other philosophical positions in the light of artificial life research (e.g., Aristotle's position on the nature of life, Matthews, 1996; Spencer and Dewey on the concepts of life and mind, Godfrey-Smith, 1996).

Additionally, philosophical debate concerning concepts central to artificial life as a whole, such as agency, autonomy, complexity, emergence, self-organisation, life-as-it-could-be, and life itself, are situated within High Class artificial life. Such debates have often predated the emergence of artificial life, and have been subject to many different styles of enquiry (e.g., on the meaning of autonomy see Boden, 1996a). However, with the advent of artificial life technologies, formalisms, and models, new kinds of thought experiment are taking place (Langton, 1996; Ray, 1994; Kauffman, 1993). In much the same way that philosophers were able to formulate new kinds of thought experiment with the help of artificial intelligence (Searle, 1980; Churchland, 1984; Clark, 1989), cellular automata, digital ecosystems, NK fitness landscapes, etc., provide arenas within which the utility of conceptual schemes may be explored.

Unfortunately, these automated thought experiments are often doing double duty, providing, on the one hand, interesting proofs of concept, and philosophically stimulating artefacts, and on the other attempting to explain things about the real world, i.e., to do science. Although their worth in terms of conceptual analysis is becoming increasingly apparent, stimulating much debate within various philosophical disciplines, and even impacting on discourses as distant from computer science as social history,[1] their utility as scientific models has yet to be demonstrated

---

[1] For example, in a recent radio discussion programme Jonathan Miller, medical doctor and self-styled Renaissance man, has appealed to the powers of self-organization in an attempt to resolve the apparent paradox presented by, on the one hand, the discovery of Hitler's indolence, and, on the other, the intensely structured order of the third Reich. Miller quoted recent evidence from artificial life suggesting that complex order may arise from simple local interactions, rather than requiring global co-ordination from some central executive.

unequivocally (di Paolo, 1996b; Noble, 1997; Miller, 1995).

A central problem for High Class artificial life concerns the definition of ‘life-as-it-could-be’. Until this notion is clearly defined (a process which will probably require the prior definition of life itself), work addressing it will be inherently philosophical. Bringing the results of such work to bear on natural phenomena (e.g., Ray’s, 1994, exploration of parasitism in a digital ecosystem; Kauffman’s, 1989, NK fitness landscape formalism as a model of the Cambrian explosion; Langton’s, 1991, characterisation of “life at the edge of chaos”) is problematic, since without a principled understanding of the relationship between artificial systems and their natural counterparts, the similarities and differences between prospective examples of life-as-it-could-be and life-as-we-know-it are hard to assess (see, for example, Jongeling, 1996, for a critique of Kauffman’s NK fitness landscape as a model of the Cambrian radiation).

For example, the similarities between Ray’s (1994) artificial ecological system, Tierra, and the natural ecosystems he has studied may merely be superficial, whereas the differences may be telling. Ray has noted that, when comparing the observed behaviour of his artificial ecologies to the known behaviour of natural ecologies, “It seems that what we see is what we know” (Ray, 1992, p.21), i.e., that natural phenomena (e.g., parasitism) may be observed within digital ecologies. However, without an established theory of life which specifies the grounds upon which comparisons between artificial and natural life may be made, the interpretation of such digital ecologies as exhibiting life-like behaviours might result from what researchers “know” influencing what they “see”, i.e., that trivial resemblances between artificial and natural systems might lead to bogus inferences from from one class of system to the other.

2.4.2 Model Class Artificial Life

Research within this class of artificial life concerns itself with the validity of scientific theory. As such it is most naturally understood as science. In much the same way that artificial intelligence provided a modelling paradigm which supported the cognitive sciences, artificial life provides a modelling paradigm capable of supporting the biological sciences.

Evolutionary simulation models clearly lie within this class of artificial life. Additional examples include computational neuroethology (Cliff, 1991; Beer, 1990), ethological experiments using robots (e.g., Webb, 1994; Franceshini, Pichon, & Blanes, 1992), experiments with virtual ecologies (e.g., Colasanti & Hunt, 1997), and models of morphogenesis (e.g., Kitano, Hamahasni, Kitazawa, Takao, & Imai, 1997).

Within this class, there is little room for purely exploratory scientific research undertaken with little notion of a theoretical framework capable of supporting it. Whilst such innovative work makes a potentially valuable contribution to the development of a scientific discipline through discovering new perspectives on old problems, uncovering new problems within existing frameworks, or exploring novel scientific tools, it must be appreciated for what it is, namely speculative, pre-theoretical inquiry. As such, although purely exploratory research may be ultimately intended to contribute directly to scientific endeavour, it is more naturally thought of as philosophy until such time as it is integrated within a scientific theoretical framework.

Scientific pursuit within artificial life will be dubbed Model Class research in an attempt to distinguish this undertaking from some kind of empirical research concerning artificial systems.

Although natural sciences have an empirical element in addition to any theoretical or modelling element, within the artificial sciences there is no straightforward equivalent to the collection of field data. Newell and Simon (1976) attempted to underwrite ‘computer science as empirical enquiry’ with their ‘Physical Symbol Systems Hypothesis’. The extent to which they were successful is debatable, but, as yet, within artificial life, no comparable framework has been forthcoming. Individual artificial life practitioners have proposed various schemes analogous to that of Newell and Simon, but currently no one position is capable of supporting empirical enquiry within artificial life.

For example, Langton (1991) has proposed an information-theoretic account of life. Langton claims that “in living systems, a dynamics of information has gained control over the dynamics of energy, which determines the behavior of most non-living systems” (Langton, 1991, p.41). Such a notion potentially offers a theoretical framework with which to assess the relationship between life-as-we-know-it and life in general, i.e., to underwrite ‘artificial life as empirical enquiry’. However, Langton’s notions are, as yet, impoverished in comparison to the framework, erected over Newell and Simon’s foundations, which attempts the same task for artificial intelligence. Furthermore, even the more established framework of artificial intelligence cannot be claimed to have unproblematically achieved this task (Dreyfus, 1979; Searle, 1980; Winograd & Flores, 1986).

Until artificial life possesses a mature framework of this kind, the only valid scientific data within artificial life will be ‘experimental data’, i.e. ‘data’ derived from models with a view to testing theory rather than exploring some virtual reality.

In contrast to this position, artificial life computer simulations are sometimes presented as worthy of investigation for their own sake, since natural phenomena have been observed within them; “Communication evolved within this world”, “Different classes of parasite evolved within this world”, “Mean fitness increased within this population when tools were introduced”. Such “digital naturalism” (Fontana, Wagner, & Buss, 1994) claims to advance understanding of the natural phenomena which are mirrored within the artificial worlds being explored through the collection of ‘empirical data’.

However, the ‘creator’ of such ‘artificial worlds’ is confused if she feels that she validly mimics the naturalist in simply observing her subject matter under various conditions. True naturalism takes place within an overarching theoretical framework, marshalling observations in order to support, or challenge, current biological theory. In contrast, the observations made of an artificial world constructed within no such framework can neither challenge, nor support, any theory with application wider than the artificial world itself. Such observations can serve no theoretician whose interests reach further than a full understanding of a specific artificial world. The extent to which the facts revealed by such observations constitute new knowledge is simply the extent to which the creator of an artificial world initially failed to understand it.

In baldly comparing and contrasting an artificial world with the real thing, the creators of such artificial worlds are attempting to both have their cake and eat it. However, there is no cake to be had in any appeal to ‘interesting’ similarities between the artificial world and the natural world, nor is there any cake to be eaten in drawing attention to ‘interesting’ contrasts between them. Unless such parallels were previously hypothesised to exist, and can be understood within some theoretical framework, they are either merely accidental (and thus not interesting), or merely purposed (and thus not interesting).

Within experimental scientific paradigms, no project is validly undertaken without an explicit hypothesis in mind; an explicit hypothesis requiring a theoretical framework within which to embed it, a reasonably rigorous vocabulary with which to express it, etc., etc. Under such a paradigm, theory precedes experiment, informing and validating experimental design. Simulation becomes a means of testing hypotheses, of exploring the consequences of theories, of revealing the implications of a scientific position. The gathering of observations ceases to be an aimless whim, becoming a process with a goal wider than merely understanding a specific simulation. For an experimental scientist, the collection of observations is not valuable in and of itself, as certain simulation designers would seem to have us believe, but is only valuable with respect to hypotheses within a theoretical framework.

Attempting to reconcile the findings of digital naturalists with those of natural naturalists is, in one sense, doomed to succeed, since the systems which digital naturalists explore are *designed to resemble natural systems*. However, the chances of these resemblances being the product of systematic, theory-embedded, parallels between the natural and the artificial are, almost certainly, vanishingly small.

Attempting to discover the nature of life (or some other biological phenomenon) through digital naturalism is analogous to attempting to discover the laws of aerodynamics which govern flight through drawing many different pictures of imaginary birds, imaginary flying insects, etc. Some pictures are of steel birds, some of birds with their beaks up-side down, or back-to-front, some of birds without wings, some of birds whose legs are detached from their bodies, etc. Once many such pictures of ‘flight-as-it-could-be’ have been rendered, one might spend years searching for principles which reveal the nature of flight. Without some theoretical framework within which to locate this enterprise, some framework which takes existing theory concerning ‘flight-as-we-know-it’ as a bedrock on which to base the drawings which, in our thought experiment, constitute ‘artificial flight’, the chances of stumbling across some of the fundamental principles of aviation are remote in the extreme.

2.4.3 Working Class Artificial Life

Research within this class of artificial life concerns itself with the production of useful artefacts. As such it is most naturally understood as engineering. The products of this class of research are not claimed to be instances of life, or to elucidate the nature of life; rather they are intended to have practical utility as a result of possessing some life-like qualities (e.g., their utility is a result of their adaptive, evolved, co-operative, self-regulating, unpredictable, etc., properties).

Research programmes within this class include evolutionary robotics (Harvey, Husbands, & Cliff, 1993), hardware evolution (Sanchez & Tomassini, 1996), evolutionary chemistry or directed molecular evolution (e.g., Bartel & Szostak, 1993), and the development of software agents (Reynolds, 1987), artificial-life-based entertainment software (e.g., Grand, Cliff, & Malhorta, in press), and artificial-life-inspired art (e.g., Sims, 1995; Hutzler, Gortais, & Drogul, 1997).

Although scientific theory will often inform the development of these products, and, reciprocally, their development may well have implications for scientific theory, from the perspective of working class artificial life such theory is a means to an end, rather than an end in itself.

This class of artificial life research has proved to be the most successful so far. Measured by its own (commercial) criteria, working class artificial life has had a considerable impact (e.g., *Creatures*, a piece of entertainment software in which users interact with life-like artificial agents, was, at the time of its release, the fastest selling game ever for PCs running Windows 95) and, if it continues to develop analogously to artificial intelligence[2] could enjoy considerable success despite failing to realise its initial grand claims (Brooks, 1991).

2.4.4 Fluidity and Constraints

Individual pieces of research within artificial life may simultaneously involve philosophical, scientific, and engineering aspects. Thus class membership is fluid, and often individual pieces of research will move rapidly between one class and another.

For example, Tim Smithers is primarily a Working Class roboticist, interested in designing robot architectures and controllers which enable, amongst other things, autonomous locomotion in uncertain environments, yet he finds it necessary to outline a conceptual space in terms of control theory and cybernetics within which he feels comfortable describing his work (Smithers, 1992). In this respect Smithers' research involves both High Class and Working Class activity. Similarly, although Webb's (1994) robot model of a cricket is a Model Class piece of research which has served to challenge previous theories of cricket phonotaxis, as it involves a physical robot, many Working Class issues had to be faced before the model could be constructed.

What differs between these classes of artificial life research are the constraints imposed upon the practitioners operating within them.

The constraints under which High Class artificial life research is produced are, at first glance, constraints of coherence, parsimony, etc. Therefore, although work within this class is sensitive to findings within the lower classes of research, it is judged on essentially internalist criteria, i.e., is it intelligible? Secondarily, the conceptual frameworks which are the product of High Class artificial life stand or fall on their ability to direct research within the lower classes of artificial life research, and to accommodate the results of such research, and perhaps research from outside artificial life. Within High Class artificial life there thus exists a dialectic between internally and externally imposed constraints.

A Model Class scientist must design a model that is constrained by a scientific theory. The model may then be used to test the theory in much the same way that an experiment may be used to test a theory. The construction of a good model, like a good experiment, involves no design decision that is not motivated by theory. Any extraneous effects must be controlled for, irrelevancies must be counterbalanced, and artefactual results must be guarded against.

In contrast a Working Class engineer is constrained by no theory (although she may be motivated by theory). Instead, reality itself constrains the design process. The design of Working Class systems must be such that they function adequately in the real world.

Thus, although individual pieces of artificial life research may involve elements from all three classes, the constraints on work within each class differ fundamentally. In order to ensure the validity of such research care must be taken to guarantee, for example, that the design of Model

---

[2] Witness the recent defeat of chess grandmaster Garry Kasparov at the 'hands' of IBM's supercomputer, Deep Blue. The artificial chess player is claimed to neither possess intelligence, nor reveal anything about intelligence, it is simply very good at something which is typically thought of as requiring intelligence.

Class systems is not constrained by reality, but by theory; that the conclusions drawn from High Class research are not considered to be scientific findings, but the results of conceptual exploration; and that the examination of Working Class artefacts is not taken to be a scientific undertaking, but an engineering one. That is, care must be taken to ensure that the correct constraints are applied to each element of the research.

Many artificial life researchers refer to the systems they design almost interchangeably as simulations, or artificial worlds, or models. The difference between the latter term and the former two terms is that to be a model an artificial system must resemble some other system systematically in some theoretically motivated ways, whereas an artificial world, or simulation, merely has to resemble the real world in more arbitrary ways — perhaps closely, perhaps hardly at all, perhaps within practical limits, perhaps within limits imposed by aesthetic, or market forces. Such systems may be constrained by internal factors (does it make sense?), or real-world factors (will it sell?). Only once such systems are constrained by theoretical factors (does it test a theory?), do they become models.

In summary, as regards evolutionary simulation modelling, the class system presented above stipulates that such a practice must be theory led. Rather than designing simulations to merely emulate natural phenomena (e.g. MacLennan, 1991; Werner & Dyer, 1991), allowing reality to constrain one's simulation design, and subsequently exploring these simulations as though they were 'artificial worlds' (i.e., digital naturalism, Fontana et al., 1994), models must be designed to capture the assumptions made by specific theories. Once such models have been designed, theories can be tested. All that can be achieved through digital naturalism is the collection of many different, detailed accounts of carefully designed artefacts (in both senses of the word).

2.5 Arguments Supporting the use of Artificial Life Simulation as a Theoretical Tool

A simulation is a model that unfolds over time. Rather than constructing static representations of the process under examination, such as flow charts or equations, and relying on human interpreters to simulate the passage of time, or determine the state of the system at some arbitrary time analytically, the simulation designer captures the dynamics of the original process by specifying dynamic mechanisms which govern how the system changes over time. The character of such a simulation's dynamics is determined experimentally, through allowing the simulation to unfold over time.

A common practice in scientific modelling involves resorting to numerical, iterative, or approximate solution techniques in situations that do not admit of analytic solutions. When an equation, or set of equations, cannot be solved analytically, an iterative algorithm is implemented that must be repeatedly called in order to determine how the system behaves. The state of the system at some arbitrary point cannot be determined in a manner other than iterating the model for the required period of time.

Recently, there have been a number of papers published within the artificial life literature advocating the application of artificial life simulations to problems within evolutionary biology (e.g., Bonabeau & Theraulaz, 1996; Ray, 1994; Taylor & Jefferson, 1994; Miller, 1995; di Paolo, 1996b; Sober, 1996; Noble, 1997). Within these papers, the primary advantage of a simulation-based approach to evolutionary modelling is identified as the fact that the design of simulations is often more tractable than the construction of more formal, analytic models (e.g., the coupled differential equations comprising game theoretic models). In addition to this claim, the authors of these papers raise many further points, and make many further comparisons in favour of simulation modelling. Some of these points and comparisons are made by many authors, some by only a few. This section attempts to organise and assess these various arguments.

2.5.1 Equations vs. Simulations

The arguments favouring the use of simulation models typically contrast this approach, which is sometimes characterised as properly situated exclusively within artificial life, with what have been dubbed ‘equational models’ (Taylor & Jefferson, 1994), which are sometimes characterised as typical of modelling approaches within evolutionary biology.

Some authors claim that these two modelling practices can be distinguished on the basis of clarity, explicitness, and inter-subjectivity. For example, Miller (1995) claims that simulations are more explicit models than those built from differential equations, since, during simulation design, the processes which govern the evolution of the system must be rendered as particular pieces of code. Miller also claims that simulations may be passed easily between researchers allowing more effective peer-validation of simulations than of equational models. Similar claims are made by Taylor and Jefferson (1994) who state that the “explicit” representation of behaviour within a simulation compares favourably with the “implicit” representation of an organism’s behaviour within equational models as “the solution to equations which must be integrated”. The authors also maintain that simulations are a more direct “encoding” of behaviour, and that this facilitates their design, use, and modification, to such an extent that these processes are necessarily easier to carry out for simulations than equational models.

These unqualified claims are misleading. For example, it is equally admissible to claim of equational models that they capture theoretical assumptions *more* explicitly than simulations since they do not involve extraneous processes which are necessary in order to implement the model as an unfolding, automated process, but which are not spoken to by the theory being tested, and are thus the source of potential artefacts. Similarly, the claim that simulations can be exchanged by modellers in order for their validity to be checked, can be made more forcefully for equational models, which can be presented in their entirety within an academic paper, rather than requiring an additional exchange of computer code.

In general, unqualified claims of the superiority of one style of modelling over another are not compelling. Clarity, ease of design, ease of presentation, etc., will vary from model to model to a greater extent than they vary from modelling style to modelling style.

2.5.2 Conservative Simulation

Rather than assert the superiority of simulation-based paradigms, some authors have characterised them as useful aids to theorists interested in problems which are poorly suited to exploration through equational models.

Sober (1996) proposes that simulations are useful in enabling theoreticians to “get a feel for the model’s dynamics” in cases for which the model proves to be unsuited to analytic methods. Similarly, Bonabeau and Theraulaz (1996), whilst advocating the widespread use of simulation techniques as useful tools for scientific enquiry, characterise such simulations as "inspiring metaphors" rather than fully functioning models.

When are simulations likely to prove more useful than more analytic approaches? Sober suggests that when computer simulation is "realistic enough and the results of computer simulation surprising enough, no one can deny the pay-off" (Sober, 1996, p.362). More explicitly, several authors (Taylor & Jefferson, 1994; Miller, 1995; di Paolo, 1996a) identify characteristics of equational models which simulation-based models improve upon, and, on the basis of these characteristics, propose a more influential role for simulation than that of mere thumb-nail sketch, namely that of a fully-fledged modelling technique with which to support of refute hypotheses.

Miller (1995), and Taylor and Jefferson (1994), state that certain mathematical assumptions which are made in the construction of tractable equational models may be relaxed under a simulation-based regime. The infinite, random-mating, unstructured populations often assumed within evolutionary models based upon differential equations may be replaced with finite, structured populations in order to highlight effects of genetic drift, frequency dependent selection, extinction, and other evolutionary phenomena. In addition, the difficulties faced by equational models featuring non-linearities, or increasingly complex inter-dependencies between the actions of agents, are largely absent from simulation-based models. Further characteristics of natural phenomena which prove difficult to incorporate within equational models include the representation of spatially distributed phenotypes, and repeated interactions between individuals (Nowak & May, 1992; Lindgren & Nordahl, 1994).

The difficulty in constructing equational models of many verbal arguments is highlighted by Miller (1995) and di Paolo (1996a). Both suggest that simulation models of such arguments might prove easier to construct.

The suggestion that simulations might profitably be employed in the augmentation of existing evolutionary modelling will be termed *conservative* simulation modelling. As Sober (1996, p.362) notes "Population geneticists have used computers in this way since the 1960s . . . I see nothing controversial about this use of computers. By their fruit shall ye know them."

2.5.3 Radical Simulation

In addition to the conservative position outlined above, a more *radical* use for simulation models has also been proposed. di Paolo (1996a) expresses this radical approach as the contention that simulations may be designed to test theories which are inexpressible within the paradigm associated with existing evolutionary models, rather than merely to reimplement formal analytic models.

There are two aspects to this radicalism. The first aspect involves the wide spread notion that simulation models are able to address emergent, or self organising phenomena as a result of their 'synthetic' nature (as opposed to the supposed reductionist nature of equational models). It is claimed that since simulations are built from low-level mechanisms (e.g., artificial organisms) which instantiate low-level behaviours (e.g., locomotion), they have the potential to explore the nature of phenomena which although not explicitly striven for by (or coded within) the low-level mechanisms, are nonetheless robust aspects of their high-level, aggregate behaviour (e.g., flocking). This intuition underlies the assertion by Miller (1995), Bonabeau and Theraulaz (1996), and Taylor and Jefferson (1994), that the strength of artificial life simulation models lies in their ability to model natural phenomena which are highly complex, and possibly emergent, or self-organising, and that it is through such modelling that artificial life simulations will prove to be most useful since these phenomena are hard to model using analytic techniques – "analytic approaches are certainly doomed" (Bonabeau & Theraulaz, 1996, p.315).

However, the danger of associating the utility of this modelling technique exclusively with a specific set of problems (e.g., the role of self-organisation in evolution), the importance of which have yet to be ascertained, seems potentially limiting. This conflation of modelling techniques with philosophical positions was addressed in section 2.4 above.

A more considered kind of radicalism is presented by di Paolo (1996a). The author suggests that simulation modelling may present an opportunity to examine the philosophical assumptions which are associated with existing evolutionary modelling techniques, and to explore their consequences. This line of enquiry will be pursued in section 2.6 below in which the perspective on evolutionary phenomena afforded by evolutionary models derived from economic game theory will be contrasted with that afforded by evolutionary simulation models.

2.5.4 Simulations as Sources of Empirical Data

Within the literature discussed above can be found one additional class of argument in favour of the construction of evolutionary simulations. Many authors (Bonabeau & Theraulaz, 1996; Ray, 1994; Taylor & Jefferson, 1994; Miller, 1995) claim that evolutionary biologists face a problem in that they possess scant evidence with which to reconstruct the evolution of life on earth. Such evidence includes the fossil record, and *in vitro* evolutionary experiments carried out in the laboratory. Compounding this claimed paucity of raw data, the authors point out that any evidence that biologists do possess can only be the result of one evolutionary sequence, since life has (presumably) only evolved once on earth. From these observations originates the claim that, with only terrestrial life to draw upon in constructing theories to explain phenomena associated with life, biologists are unable to distinguish between aspects of life which are contingent upon the particular historical development of life on earth, and those aspects which are fundamental to life in general.

Artificial life simulations are presented as an alternative source of empirical data with which to settle disputes over the contingent vs. fundamental status of characteristics of living systems. The collection of data from simulations offers many advantages over the collection of data through traditional field experiments, or naturalistic observation. Simulations are repeatable, and manipulable; they are easily observed, offering complete access to relevant variables, and involving no danger of observer presence interfering with observed behaviour. In addition, elementary statistical analysis of the data which simulations provide may be automatically carried out by the simulations themselves.

Although all these characteristics may be useful properties of simulation models, they fail to adequately compensate for the fact that empirical data cannot be gathered from simulations in order to augment that collected from the natural world (see section 2.4.2 above). A simulation model which purports to offer new empirical data concerning, for example, the Cambrian explosion, or an 'artificial radiation' which is claimed to be similar to the Cambrian explosion in some important respects, cannot be claimed to be providing *empirical* data in the same way that the fossil record does. It is merely a model of a major speciation event which, if competently constructed, might make some predictions about speciation, the creation of niches, or associated phenomena, which could be falsified by subsequent empirical discoveries. In this respect, such a simulation model is no different to an equational model of the same phenomenon (Noble, 1997, provides a similar argument).

2.5.5 Problems with the Simulation Approach

In a nutshell, the primary problem with simulation modelling is its relative under-development. Whereas equational models benefit from the maturity of differential calculus and associated mathematical tools used in their construction and analysis, simulation models suffer from a lack of such tools, and the relatively underdeveloped state of those tools which do exist (e.g., statistics, cluster analysis, dynamical systems theory). In addition, whereas the mathematical skills required by theorists are taught as a matter of course throughout the natural sciences, the analogous skills of use in the design of simulations are not taught as a unified whole. For the most part such skills are poorly understood by modellers, and must typically be developed independently by each. Furthermore, typical simulation components such as genetic algorithms, artificial neural networks, cellular automata, etc., are themselves, as yet, not fully understood, and their application to scientific problems is in its infancy.

This lack of a coherent, unified body of practical knowledge with which to guide the design and analysis of simulation models typically manifests itself as an inability to construct simulations which adequately guard against artefactual results, and an associated inability to thoroughly account for the observed behaviour of simulations.

For example, compare the results of an imaginary equational model with those of a simulation model of the same evolutionary phenomenon. The solution to a set of differential equations which constitute an equational model of some evolutionary system might describe the conditions under which a particular class of equilibria exist. For example, the inequality $a < b$ might define a class of systems in which a certain strategy is evolutionarily stable. This solution is informative. It captures a relationship between two variables, a and b. Presumably these variables have some understandable role to play within the model as a whole, perhaps a stands for ‘weight of female at onset of oestrus’, whilst b stands for ‘weight of male at onset of female oestrus’.

In contrast, a simulation designed to explore the same theory as the equational model described above, might result in several sets of data points which describe some aspects of the system as it varies over evolutionary time, and over the space of possible parameterisations of the model, and with the initial conditions from which the system evolves, etc. Upon analysis, these data points should reveal the same relationship as that discovered through the construction and analysis of the equational model, but the move from ‘raw solution’ (data points) to ‘informative solution’ (relationships) is more involved than the straightforward process described for the equational case outlined above.

Some effort must be made to reconstruct the *relationships between classes* that mathematical models utilise in explaining the behaviour of analytically derived models, from the *instances* which the simulation model generates.

Thus, although, under certain conditions (see section 2.5 above), the construction of simulation models might prove more tractable than the construction of analogous equational models, the analysis of such simulation models often requires an additional effort which threatens to more than compensate for any increased ease of design. This situation has been dubbed the ‘law of uphill analysis and downhill invention’ (Braitenberg, 1984), and has been given a more thorough exposition by Clark (1990), in terms of a distinction between automatic design processes and manual design processes.

Clark argues that manually designed systems are easy to interpret because the processes involved in their creation provide a natural way of decomposing them into intelligible sub-systems. For example, a machine manually designed to catch a thrown tennis ball might comprise various sub-systems charged with effectively carrying out part of the ball-catching problem. Upon completion of the ball-catching system, an analysis of the manner in which it achieves its task is trivial (for the system’s designer) since the manual design process involved specifying precisely this.

In contrast, a system designed by an engineer using an automatic design process (e.g., harnessing some kind of artificial evolutionary process in order to design the tennis-ball catching system) is not so amenable to analysis. Although the engineer specified the problem, and the resources upon which the solution might draw (e.g., the general architecture of the system), she had no say in the manner in which the automatic process saw fit to exploit these resources, i.e., she had no part in constructing the algorithm which the automatic design process settled upon.

In the language of Peacocke’s (1986) extension to Marr’s (1977) characterisation of design and explanation within artificial intelligence, the engineer employing an automatic design process has no input beyond level 1.5 of the three-stage “classical cascade” which design and explanation typically involve.

Within Marr’s original account, level 1 of his three-stage hierarchy consists of specifying the problem (e.g., the system must be able to catch balls of a particular size, thrown from a particular range of positions, within a particular range of velocities, etc.). Marr termed this the *computational* level since, in order to be solved through the design of some information processing system, the problem must be expressed in a language amenable to implementation as a computer program.

Level 2 involves specifying an *algorithm* capable of solving the problem (e.g., one algorithm might calculate the future trajectory of the ball using Newtonian mechanics, and on the basis of the ball’s projected flight, calculate an intercept trajectory for the catching limb). Marr (1977) gives the example of Fourier analysis which may be calculated using any one of a number of algorithms.

Finally, level 3 is an *implementation* of this algorithm (e.g, a particular LISP program, or a particular circuit board, or a particular system of mechanical devices). Marr’s system of levels is hierarchical since many level-3 implementations might capture any individual level-2 algorithm. and, likewise, many level-2 algorithms might solve any individual level-1 problem.

The account of Marr’s system offered here concentrates exclusively on problems which have what he terms a Type-1 theory. Such problems can be solved in a manner which allows insights into the nature of the problem to be gained, i.e., Type-1 theories are explanations. For example, in solving a problem presented by the pattern of inheritance evidenced by peas, Mendel achieved some insight into the particulate nature of genetics. In doing so Mendel moved towards an *explanation* of inheritance, rather than a mere *description* of it.

In contrast, problems which only have a Type-2 theory may perhaps be solved, but such solutions will not throw light upon the nature of the problem, i.e., Type-2 theories are descriptions. Marr offers protein folding as an example of a problem which only has a Type-2 solution, since this problem "is solved by the simultaneous action of a considerable number of processes, *whose interaction is its own simplest description*" (p.134, Marr's italics). I will ignore the possibility that artificial life may be addressing problems which only admit of Type-2 theories, at least for the moment, since such theories cannot underwrite explanations of problems, but merely descriptions of them. [3]

As an addition to Marr's original scheme, Peacocke's level 1.5 involves specifying slightly more than a description of the problem, but slightly less than a full solution algorithm. At level 1.5, the designer must decide upon what Clark (1989) terms a *competence theory,* by which is meant a characterisation of the problem which is "more than merely *descriptive*" of the solution to be discovered, but is also "*suggestive* of the processing structure of a class of mechanisms" within which the solution is to be searched for (Clark, 1989, p.285). For Clark (1989, p.285), "A competence theory, then, leads a double life. It both specifies the function to be computed, *and* it specifies the body of knowledge or information which is used by some class of algorithms." Level-1.5 descriptions are thus intended to specify the range of resources upon which the solution may draw (e.g., the space of mechanical devices from within which the tennis-ball catcher must be selected).

For the manual designer, there is little difference between level 1.5 and level 2. The resources upon which the solution algorithm may draw are intimately connected with the design of the solution algorithm itself. However, for those engineers employing automatic design processes (e.g., genetic algorithms, or artificial neural network learning algorithms), level 1.5 is as far down the classical cascade as it is necessary to descend.

By this stage an engineer employing an artificial neural network learning algorithm will have decided upon the number of nodes in her network, the class of learning algorithm to be employed, the format of the input to, and output from, the network, etc.; an engineer employing a genetic algorithm will have specified a mapping from genotype to phenotype, and settled upon styles of fitness appraisal, reproduction, etc. These preliminary machinations specify the resources upon which the solution may draw; they define a space of possible solutions which, it is hoped, (a) contains a viable solution, and (b) is not so complex as to be overly expensive to search.

Once this has been accomplished the system will proceed, upon execution, to automatically generate a solution. The modeller has thus moved from level 1.5 directly to level 3, and is in possession of a working implementation (e.g., a system which can catch tennis balls).

For the engineer, this by-passing of level 2, which is where most of the design work takes place, is a boon. For the scientist in search of an explanation of how the system works, it is a mixed blessing as the manual move from level 1 to level 2 is also where most of the insight that fuels explanation is gained. The scientist in possession of an automatically designed model is only half-way home. She must proceed to work backwards from her level-3 implementation to a level-2 understanding of how the system actually works (e.g., faced with a working tennis-ball catcher which is implemented as an artificial neural network, effort is required in order to discover how

---

[3] Thanks to Mike Wheeler for alerting me to this complication.

the network achieves the task).

This characterisation of the difference between manual and automatic design offered by Clark (1990), in his terms a “methodological inversion”, is reminiscent of the ‘no free lunch’ maxim. What one gains in ease of design, one loses in ease of explanation.

For our current purposes, the argument seems equally applicable to a characterisation of the difference between equational and simulation modelling. Although simulations may sometimes be easier to build than equational models, they require more effort in analysis than their analytic cousins.

Under this reading, at level 1 both equational and simulation modellers must characterise their hypothesis in an adequately rigorous manner. Each modeller then proceeds to outline the style of hypothesis testing she envisages. Here the two modellers part company. For the equational modeller, this level-1.5 account is in terms of the character of the relevant variables and their interrelations, whilst, for the simulation builder, it is in terms of the mechanisms governing system evolution.

The equational modeller then proceeds to struggle with her system of differential equations, until, through cunning and scholarly hard work, she arrives at a solution. In contrast, her simulation building companion merely unleashes her automatic simulation process, and waits for it to collate the (typically massive) set of data points which constitute her solution. However, upon its discovery, the equational solution is straightforwardly interpretable in terms of its components, consisting as it does of intelligible relationships between meaningful sub-parts. In contrast, the set of data-points facing the simulation-based modeller is far from straightforward to analyse, requiring considerable scholarly effort in the form of statistical analysis, etc., before the implications of the data for the theory being explored are clear.

The difficulties involved in such analysis can be overcome in two ways. Keeping simulations simple ensures that the analytic tools available can be applied usefully. Ensuring that simulations are models of theories ensures that, during analysis, the theories can be used to unpick the simulations.

Much of the research which has repeatedly been distinguished from evolutionary simulation modelling within this chapter (e.g. MacLennan, 1991; Werner & Dyer, 1991) can be characterised as modelling work which proceeds from a poorly specified level-1 problem (e.g., can communication evolve in a particular artificial world?), to a well specified, but atheoretical level-1.5 specification of the resources upon which an automatic process may draw (e.g., a certain genetic algorithm). From this competence level description, the modellers move directly to a level-3 implementation of a system which exhibits behaviour with superficial similarities to natural phenomena (e.g., communication). What is missing from this research is a move from this level-3 system to a level-2 understanding of the implications of the model for the theory it was testing.

The reason this final move is missing is because, lacking a hypothesis (within a theoretical framework) to test, such research has no need for it. Unfortunately, this lack of theoretical rigour is typically disguised by claims that the problem has, in Marr’s terminology, no Type-1 theory, but merely admits of Type-2 theories, by which I mean that researchers are often satisfied with the raw solution to a problem, an existence proof, believing that a more principled understanding of the nature of the problem is made impossible by its emergent, chaotic, non-linear, or context sensitive nature. Although some problems no doubt are of this type, merely assuming that this is the nature of the beast is an unproductive strategy.

Marr has this to say on the subject (whilst he is addressing research within artificial intelligence, his comments seem equally applicable to research within artificial life)...

> "The principal difficulty in artificial intelligence is that one can never be quite sure whether a problem has a Type-1 theory. If one is found, well and good; but failure to find one does not mean that it does not exist. Most artificial intelligence programs have hitherto amounted to Type-2 theories, and the danger with such theories is that they bury crucial decisions, that in the end provide the key to correct Type-1 decompositions of the problem, beneath the mound of small administrative decisions that are inevitable whenever a concrete program is designed. This phenomenon makes research in artificial intelligence difficult to pursue and difficult to judge. ...
>
> With any candidate for a Type-2 theory, much greater importance is attached to the performance of the program. Since its only possible virtue might be that it works, it is interesting only if it does. Often a piece of artificial intelligence research has resulted in a large program without much of a theory, which commits it to a Type-2 result, but that program either performs too poorly to be impressive or (worse still) has not even been implemented. Such pieces of research have to be judged very harshly, because their lasting contribution is negligible." (Marr, 1977, p.135).

2.6 Evolutionary Simulation Models and Game Theoretic Models

In this section the assumption that evolutionary simulation modelling does not belong within evolutionary biology, an assumption under which this chapter has been labouring until now, is discarded. A brief account of simulation modelling within evolutionary biology is offered, and differences between this practice and evolutionary simulation modelling as it is conceived of within this thesis are discussed.

In principle a modelling paradigm with the flavour of evolutionary simulation modelling, as it has so far been described, could be located at the fringes of evolutionary modelling within biology, or game-theoretic modelling within economics, rather than within artificial life. Indeed, as this chapter has unfolded it has become apparent that the work which I am describing has a great deal in common with the simulation work undertaken by some game theorists. More specifically, such work is redolent of "replicator dynamics" studies within economic game theory (Binmore, 1991), and simulation work used to support the findings of game theoretic models of evolutionary or developmental dynamics within biology (e.g., Maynard Smith, 1982, Chapter 5, "Learning the ESS").

However, although there are many similarities between evolutionary simulation modelling and these rather recent developments within various branches of game theory, the style of modelling presented here has fundamental differences from game theoretic models as they are typically utilised.

These differences can be crudely characterised by distinguishing between game theoretic models and evolutionary simulation models on the basis of the perspective each takes towards the dynamics of adaptation. Whereas evolutionary simulation models, and artificial life in general, are concerned with dynamic change, game theoretic models are principally concerned with stasis.

2.6.1 A Dynamic Perspective on Adaptive Systems

Artificial life is almost exclusively concerned with systems which change over time, and, furthermore, the *manner* in which such systems change over time. Whether the system be a cellular automaton "game of life", an autonomous robot designed to navigate an extra-terrestrial terrain, or a model of "complexity at the edge of chaos", it is considered as a time-varying system with a certain dynamic character. It is this character which is of interest to the artificial life practitioner. Will the dynamic character of the cellular automaton admit of "universal computation"? Will the dynamic character of the autonomous robot result in robust walking behaviour? Will the dynamic character of the model exhibit "complexity at the edge of chaos"?

Despite being similarly concerned with systems which change over time (economies, individual economic agents, ecologies, populations of creatures, individual creatures, etc.), game theoretic accounts of adaptive phenomena typically assume that the systems under consideration are at, or near, equilibria. Once this assumption is in place, the game theorist is faced with the task of specifying a model which admits of stable equilibria with a character which matches that of the observed system's behaviour.

For example, neo-classical economic theory asserts that since the economic agent, *homo economicus*, is an ideal maximiser of expected utility, such agents will clear a market at the equilibrium price. There is thus no sense in asking what behaviour would result from a system comprised of agents who cannot maximise expected utility. Such a system is far from equilibrium, and thus not likely to be found to reflect real economic situations in which markets are either at or near equilibria. Since, from this perspective, economic agents are assumed to be optimal players, one merely needs to identify the equilibrium price analytically in order to describe the behaviour of the market, as this is the price which will be settled upon. Attention to the global dynamics of the model is not necessary since the system will not spend time far from equilibrium.

Whilst these assumptions are, for the most part, eminently reasonable, there are scenarios in which the conditions in which real economic agents find themselves result in their inability to find the equilibrium price, e.g., sellers at an auction who are wary of the existence of cartels amongst their fellow bidders. In order to model these systems, approaches which take into account a richer pallet of dynamic behaviour are necessary (see e.g., Binmore, 1987, 1988, for critiques of the traditional axiomatic approach to game-theoretic economic modelling).

A similar perspective is evident within game theoretic accounts of evolution. Maynard Smith identifies this problem at the outset of his book, *Evolution and the Theory of Games*,

> "An obvious weakness of the game-theoretic approach to evolution is that it places great emphasis on equilibrium states, whereas evolution is a process of continuous, or at least periodic change. The same criticism can be levelled at the emphasis on equilibria in population genetics. It is of course mathematically easier to analyse equilibria than trajectories of change" (Maynard Smith, 1982, 8).

Although it is perhaps even more reasonable to expect the natural systems we see around us to be at stable equilibria given the evolutionary timescales involved, there are situations in which evolutionary systems may consistently fail to reach equilibria, or in which the equilibria which evolutionary systems do reach are more complicated than point attractors. For example, Maynard Smith follows the passage quoted above with the prediction that *cyclic* attractors will be discovered to characterise much of the behaviour exhibited by players involved in asymmetric games. This prediction has been vindicated by the discovery of a species of lizard which occurs in three distinct morphs, each of which dominates one other morph, and is dominated by the remaining morph. Such a system is analogous to the parlour game scissors-paper-stone, in which playing one move consistently will never be a lasting strategy since any such strategy can be defeated (Sinervo & Lively, 1996; Maynard Smith, 1996).

In addition, there are reasons to believe that the “trajectories of change” might be useful in distinguishing between theories attempting to account for evolutionary phenomena, even if these phenomena are characterised as equilibria within a system.

As has been pointed out by Ryan (1990), differing evolutionary theories often make the same predictions concerning current states of affairs. This is because current states of affairs are reflected in the existing data which the theories attempt to account for. However, differing theories may make differing predictions concerning the prior states of affairs which lead to these current states of affairs.

Ryan (1990) attempts to distinguish between theories which compete to account for the character of sensory systems and signalling behaviour extant in the natural world, by constructing phylogenetic histories for several species of frog. From this hypothetical history of speciation events Ryan attempts to discount certain theories whose predictions do not match the historical account he has constructed.

Evolutionary simulation modelling can contribute to this style of hypothesis testing in a way in which modelling methodologies which exclusively attend to equilibria cannot. An evolutionary simulation model provides an account of not only the behaviour of a system at equilibrium, but also the behaviour which that system passed through before it reached this equilibrium. Such accounts of the trajectories followed by evolving populations prior to (potentially) achieving equilibria might be used to distinguish between competing theories.

This is not to say that data from simulations will simply augment data from phylogenetic reconstructions (see section 2.5.4 for a discussion of this fallacious position), but that the implications of evolutionary theories for the character of evolutionary trajectories, might be clarified through evolutionary simulation modelling. The predictions resulting from such a clarification could then be compared to empirical data in the usual manner.

Despite both the predicted existence of scenarios in which the behaviour of the system under consideration is inherently dynamic, and the acknowledgement that attention to the evolutionary dynamics predicted by evolutionary theories might enable theorists to distinguish between them, game theorists typically eschew the study of dynamic change. This eschewal accounts for the accent placed on the fixed points of a model (whether they be evolutionary stable strategies in biology, or Nash equilibria in economics), rather than the general dynamic behaviour of such models. This is of course not to say that game theory cannot tolerate limit sets of a higher order than constant trajectories, or has no consideration of initial conditions or transient behaviours. However, such matters are regarded within game theory as special cases which require additional analytic techniques if they are to be addressed at all (e.g., Maynard Smith, 1982, devotes an appendix to dealing with cyclic trajectories).

In contrast, evolutionary simulation modelling will principally concern itself with the *charac-*

ter of a model's evolutionary dynamics rather than some "end product" of these dynamics, whether it be within a population of learning economic agents, a population of evolving creatures, or a population supporting a developing culture or language. From this inherently dynamic perspective, cyclic limit sets and start-up transients, are on an equal footing with game theory's cardinal limit set, the point attractor.

2.6.2 Biological Simulation

The equilibrium-centred perspective which dominates analytic modelling within biology and economics also influences the relatively infrequent use made of simulation models by game theorists. Such models, when they *are* constructed, are typically iterative numerical models used to solve differential equations which prove intractable under analytic techniques in an effort to discover the equilibria within a model, or to verify that equilibria suggested by analytic treatment of the model do indeed exist. As was mentioned above, biologists have used computers in this way for some considerable time.

For example, Iwasa, Pomiankowski, and Nee (1991) construct an equational model of the handicap principle (see chapter 5 herein) and demonstrate the existence and character of various equilibria using a simple iterative simulation to solve the difference equations which govern the manner in which their system evolves. Similarly, many examples of simulation models used to explore the evolution of learning algorithms are presented in Maynard Smith (1982, p.54–67).

An earlier model of animal conflict presented by Maynard Smith and Price (1973) was similarly implemented as a computer simulation. A conflict situation was modelled as a series of alternate moves on the part of two opponents. Each opponent was modelled in terms of a strategy which governed the probabilities with which it played each of three moves, a conventional move (e.g., a threat display involving little danger), a dangerous move (e.g., an attack likely to injure the attacker's opponent), or a retreat (at which point the confrontation ceases). For certain strategies, these probabilities were dependent on the move played by the opponent on the previous move.

For example, the Retaliator strategy starts by playing a conventional move, and will continue to play such a move for as long as its opponent similarly plays a conventional move (after a prespecified number of such moves have been made Retaliators will retreat). Should it's opponent play a dangerous move at any point during the confrontation, there is a high probability that Retaliator's will respond with a similar dangerous move. A Retaliator strategist thus bears a strong similarity to a tit-for-tat strategist within the prisoner's dilemma game (Axelrod & Hamilton, 1981; Axelrod, 1984).

Five such strategies were defined and each strategy's performance against itself, and the other four, was assessed through simulating 2000 confrontations between opponents playing the relevant strategies. The resulting mean fitness scores for each strategy vs. each other strategy demonstrated that "limited war" strategies (i.e., strategies in which conventional moves were employed) were evolutionarily stable, but total war strategies (i.e., strategies in which no conventional moves were employed) were not evolutionarily stable.

Although this model involves computer simulation and addresses the evolutionary stability of different classes of behaviour, it is not an evolutionary simulation model. In commentary on this model, Gale and Eaves (1975) point out that although the limited war ESS identified by Maynard

Smith and Price (1973) exists, an alternative ESS exists in the form of a frequency dependent mixture of total war strategists, and further demonstrate using a very simple evolutionary simulation mode, that from some initial conditions, total war equilibria, (rather than limited war equilibria) will be reached by a population of opponents reproducing asexually at rates proportional to their average weighted payoffs. The initial conditions which lead to total war include situations in which limited war strategists are rare, and situations in which the vast majority of the population never play dangerous moves.

These findings involve the evolutionary trajectories, or start-up transients, of the model, rather than any results solely concerning the potential end-points of such evolutionary trajectories (i.e., evolutionary equilibria).

More recent evolutionary simulation models include Krakauer and Johnston's (1995) model of the evolution of honest signalling. The authors apply an evolutionary algorithm to a population of artificial neural networks, which are employed as models of signallers and receivers playing a simple communication game. They pay considerable attention to the evolutionary dynamics of their system. For example, they identify oscillations of signaller and receiver fitness which are in antiphase, and interpret this as evidence of an evolutionary arms race featuring a "'Red Queen' dynamic" (p.358), in which gains made by one population are made at the expense of the other, and are necessarily shortlived.

Additional evolutionary simulation models include a series of papers by Arak and Enquist which explore the evolution of sensory biases (e.g. Enquist & Arak, 1994, and chapters 5 and 6 herein). No doubt many other examples of this style of modelling may be found within the evolutionary biology literature (see Fraser, 1957a, 1957b; Baricelli, 1957, for early examples of the application of evolutionary algorithms to biological phenomena).

Recently, models equivalent to those dubbed herein evolutionary simulation models, have been promoted within ecology, where they are termed individual-based models, or I-state models (Huston, DeAngelis, & Post, 1988; DeAngelis & Gross, 1992).

McCauley, Wilson, and de Roos (1993) provide an example of this style of ecological modelling, applying an individual-based modelling methodology to the dynamics of predator prey interactions in a spatially structured population. They construct a model in which individual agents are governed by rules for growth, reproduction, mortality, etc., and demonstrate first that the model is in agreement with an equational model, or what they term an ordinary differential equation (ODE) model, of predator-prey dynamics when populations are considered to be spatially homogeneous.

Subsequently, the authors investigate the effects of introducing heterogeneity into the population in the form of spatial distribution, age structure, etc. Within this more complex model, the conditions governing the stability of various foraging and predatory strategies (e.g., sit-and-wait, or cruising) are explored.

A precursor to this style of modelling can be found in Maynard Smith and Reichert's (1984) use of a simulation model to explore the behaviour predicted by a particular model of spider antagonism. This model allows predictions made by the model to be compared to data collected from real spiders through instantiating the behavioural model as an iterative computer program which simulated individual spiders engaging in antagonistic bouts.

Interestingly, within the burgeoning literature addressing I-state models, a debate concerning the utility of this modelling technique has emerged which in many ways parallels the issues discussed within this chapter. Whilst some authors have claimed that these models will "unify ecological theory" (Huston et al., 1988), others are more skeptical (Metz & de Roos, 1992; Murdoch, McCauley, Nisbet, Gurney, & de Roos, 1992), pointing to the lack of scientific generality possessed by the models, and raising the possibility that conclusions drawn from individual-based models might be the result of artefacts, rather than valid underlying principles.

Issues which divide the I-state modelling community include the ability of individual-based models to substantiate general scientific hypotheses rather than merely accurately simulate specific ecologies. Judson (1994b, 1994a), in her review of the development of individual-based modelling, quotes Maynard Smith's (1974a) position regarding the utility of ecological simulations. Maynard Smith maintains that the difference between models and simulations is that models strive for a minimum of detail, whereas simulations strive for a maximum. This Dichotomy frustrates I-state modellers who wish to use simulations *as* models, and thus demand a degree of generality from their simulations.

An additional bone of contention within the individual-based modelling community concerns the degree to which this modelling methodology is wedded to philosophical positions regarding the nature of the systems being modelled. Whereas Judson (1994) asserts that, amongst ecologists, a growing awareness that ecosystems may be chaotic systems has lead directly to the adoption of I-state modelling practices which are more able to capture the complexities of such systems, James Bullock (1994) has responded by suggesting that accepting the utility of these modelling techniques is in no way dependent on accepting this perspective on ecosystems. He suggests that individual-based modelling can profitably address a much wider range of phenomena which are typically addressed by more traditional modelling techniques.

From the perspective of this chapter, the first of these disputes in directly analogous to the issues raised in section 2.4 concerning the role of simulations as modelling tools, whilst the second debate is similarly analogous to the problems identified throughout this chapter in distinguishing evolutionary simulation modelling as a methodology, from the philosophical claims concerning the nature of life, etc., which often accompany it.

2.6.3 The Role of Evolutionary Simulation Modelling

The preceding argument has attempted to present a compelling case for the utility of evolutionary simulation modelling within the biological sciences. This case is made on the basis of two points. The first is that an equilibrium-centred perspective is dominant within game-theoretic, and population genetic modelling. The second is that there are reasons to believe that such a perspective precludes the appreciation of several important evolutionary phenomena. These points, taken in tandem, justify a role for evolutionary simulation modelling as an additional theoretical tool to those already in the possession of theoretical biologists. Evolutionary simulation modelling should thus be regarded as augmenting rather than replacing, or challenging, orthodox modelling techniques, enabling systematic modelling approaches to several interesting areas within biology to be undertaken.

2.7 Summary

Throughout this chapter, two distinctions have repeatedly been made, and, relatedly, two comparisons. First, evolutionary simulation modelling has been *distinguished* from much evolutionary simulation research within artificial life on grounds of scientific rigour. However, evolutionary simulation modelling has been *allied* to much of this work on the grounds of a shared perspective on evolutionary systems as primarily *dynamic* systems. Second, evolutionary simulation modelling has been *paralleled* with the rigorous evolutionary modelling practice within evolutionary biology, but *discriminated* from this practice on the grounds of a difference of perspective with regard to the dynamic character of evolutionary modelling.

Furthermore, care has been taken to distinguish this dynamic perspective on evolutionary modelling from the immature philosophical positions to which it is often affiliated within artificial life. Similar care has been taken to distinguish between the aspects of theoretical biology which result in its scientific rigour, and those which result in its adoption of an equilibrium-centred perspective on evolutionary modelling.

With these distinctions in hand, the spectrum of subject matter which may be validly addressed by evolutionary simulation modelling may be defined. The following chapter will consider this matter.

Chapter 3

The Evolutionary Perspective

> "Without the methods of Copernicus, Kepler, Galileo, and their successors in astronomy, physics, and chemistry, Darwin would have been helpless in the organic sciences. But prior to Darwin the impact of the new scientific method upon, life, mind, and politics, had been arrested, because between these ideal or moral interests and the inorganic world intervened the kingdom of plants and animals. The gates of the garden of life were barred to the new ideas; and only through this garden was there access to mind and politics. The influence of Darwin upon philosophy resides in his having conquered the phenomena of life for the principle of transition, and thereby freed the new logic for application to mind and morals and life", (Dewey, 1909, 24–25).

Possibly in allusion to Dewey's (1909) above-quoted essay, *The Influence of Darwinism on Philosophy,* Freud famously described the theory of evolution by natural selection as the second of three strikes against man's anthropocentric view of existence. Freud identified the first of these strikes as the Copernican revolution, and, modestly, the third as his own theory of the unconscious. Under this rain of blows, Freud maintained, humanity had relinquished its claims to centrality within the universe, supremacy within nature, and, finally, authority over the human mind and body themselves.

This chapter will explore the possibility that the Darwinian revolution did not just intervene chronologically between physical and psychological revolutions, but, as Dewey suggests above, that it intervenes logically between them.

Does Darwinian thought have implications beyond the biological sciences? The notion that an evolutionary perspective might be applied to areas of inquiry within the social and human sciences, as well as aspects of the life sciences, and potentially some of the physical sciences, is seductive, but politically charged. This notion has attracted many advocates, both recently and historically, but has often been accompanied by controversy. Attempts to address psychological phenomena from an evolutionary perspective have been particularly contentious, inviting accusations of prejudice and bigotry (see Lewontin, Rose, & Kamin, 1984, for a review).

Before addressing the possibility of using evolutionary simulation models to address theories external to biology, this chapter will attempt to justify the application of evolutionary frameworks outside the biological sciences by arguing that the evolutionary perspective, when adopted legitimately, stands in direct opposition to the notions of a fixed natural order which underly the positions of authors who attempt to utilise biological determinism and Darwinian theory as a means of justifying racial, or sexual, intolerance (e.g., Herrnstein & Murray, 1994).

Two themes which have impacted on the attraction of the evolutionary perspective will be explored in order to assess the validity of the application of evolutionary perspectives beyond biology. The first theme addresses the need to provide non-teleological accounts of natural systems which, although appearing to demand explanation in terms of purpose, intention and design, might more profitably be described in terms of contingent historical development under some adaptive pressure. This theme stresses the utility of the evolutionary perspective in providing such accounts.

The second theme addresses the presence of teleology within the evolutionary perspective itself. Might the language and imagery of evolutionary theory, or even the mechanisms of evolution themselves, be inherently teleological? If this were found to be the case, might not the adoption of the evolutionary perspective outside biology typically result from the attractiveness of teleological accounts, rather than the utility of the Darwinian method.

Once these two themes have been identified and assessed, evolutionary accounts within various disciplines will be considered, before a definition of the evolutionary perspective which delimits the range of subject matter that may be legitimately explored through evolutionary simulation modelling is offered.

3.1 The Potential for an Evolutionary Perspective within Scientific Enquiry

Dewey (1909) provides a fleshing-out of the three blows identified by Freud, through an account of what he terms a “transfer of interest from the permanent to the changing” (p.24). For Dewey, the Darwinian revolution was more than a biological revolution, it was the leading edge of a slow wave of philosophical change which had rolled over the physical sciences during the Renaissance and the Enlightenment, and, as a result of evolutionary thinking, had proceeded to similarly sweep across the biological sciences during the last half of the nineteenth century. Moreover, Dewey, at the outset of the twentieth century, could envisage this same movement engulfing the social sciences, and lapping at the toes of religious and moral thought.

The philosophical position which fell beneath this rising tide of dynamic thought was that of the ancient Greeks. Platonic and Aristotelian notions of a natural order which was harmonious, resisted change, and was under-written by a metaphysic which appealed to purpose, intention, and design was gradually replaced by an altogether different notion of nature as an inherently meaningless system; undirected, yet law governed.

It is in this respect that the quoted passage from Dewey with which this chapter opens likens the Darwinian revolution to the Copernican — both struck against a philosophy which reified a transcendental, fixed natural order.

Copernicus’ heliocentric model denied the divinely ordained heavenly order promoted by the orthodox geocentric model. Subsequent development of the model by, for example, Kepler complicated Copernicus’ model in ways which increasingly seemed to deny that a creator’s hand could be seen to have been at work in the organisation of the heavenly bodies. For example, celestial motion had, prior to Kepler, always been conceived of as circular, since the circle was “the most perfect form, and hence the only shape suitable for the motions of the heavenly bodies” (Wertheim, 1997, 74). Kepler’s eventual rejection of circular orbits in favour of elliptical ones, despite his own firm conviction that “the work of such a perfect Creator should be of the greatest beauty” (Wertheim, 1997, 71), represents a triumph of empiricism over dogma. However, Kepler could only reconcile his empirical findings with his religious convictions by maintaining that the elliptical forms which orbits took were poor approximations to the perfect circular trajectories which should ideally obtain. That they were only approximations resulted from the impure influence of the material world in which they resided; i.e., they were, like the shadows on Plato’s cave wall, merely reflections of more perfect forms to which they were tied by (divine) purpose and intention.

For Dewey, however, despite the inability of Kepler and his contemporaries to completely shrug off the overcoat of teleology and don the smoking jacket of materialism, this and related empirical work revealed the first cracks in a philosophy stretching back to the ancient Greeks, and was eventually to succeed in banishing purpose and intention from the physical world. With the publication of Descartes’ *Discourse on Method* came a direct expression of the contention that no characteristic of a physical system could be legitimately explained through an appeal to the system’s possession of purposive spiritual tendencies which engendered a striving toward some pre-specified aim, or *telos*, whether such tendencies were “real essences, hidden forms [or] occult faculties” (Dewey, 1909, 24).

As Dewey notes, this revolution came as “something of a shock” to astronomy, physics, and chemistry, but was prevented from sweeping all before it by the almost palpable fidelity of the argument from design as applied to the realm of living things. Whereas stars and sand might be subjugate to the impersonal laws of motion, starfish and sandworms, lowly though they were, could still be considered (ignoring the mechanistic explanations provided by Descartes) to exemplify the notion of *species* developed by the Greeks.

Under an Aristotelean philosophy which had reigned for the two millennia prior to Darwin, a species was a guiding hand (an invisible hand in the sense of Adam Smith), which governed many imperfect manifestations of a true type, holding them in line, maintaining their uniformity of appearance, behaviour and function. This guidance was not material, but transcendental, ensuring that instances of these types resisted any “disturbing forces”, in order that they achieve some final purposed end. For example, acorns distant from one another in space and time, and thus physically unconnected, each endure wind and cold, yet obstinately each develops into an oak, save those that are thwarted by some interfering influence.

Under such a reading of the concept of species, an account of their *origin*, of the manner in which species *adapt to their environment* over evolutionary time, is tautological. Species, as transcendental “first and final causes”, influencing nature without fear of reproach, do not change as the result of some materialistic process.

As Dewey recounts,

> “The Darwinian principle of natural selection cut straight under this philosophy. If all organic adaptations are due simply to constant variation and the elimination of those variations which are harmful in the struggle for existence that is brought about by excessive reproduction, there is no call for a prior intelligent causal force to plan or preordain them”, (*ibid*, p.26).

For Dewey, writing at the turn of this century, the success of this “new logic”, within first the physical, and then biological sciences, boded well for a change in perspective throughout the philosophies of mind, politics, and religion. This change in perspective would involve a rejection of the notion of a natural order in which stasis and fixity were elevated to the status of fundamental principles; in which change is regarded as a mere perturbing factor, resisted by legitimate predispositions which maintain true form. In its place would be fashioned a notion of natural order in which change was fundamental, non-teleological, and law-governed, moving from antecedents to consequents in a manner sensitive to "specific conditions of generation" (p.30). That is, a philosophy based on the notion of an ultimately directionless, yet fathomable, process, rather than notions of *species* and *telos*.

Veblen (1898) , in his paper *Why is economics not an evolutionary science?*, reaches similar conclusions to those of Dewey. Veblen points out that biologists had benefited from dispensing with a teleological notion of adaptation. However, in critically describing the practice of contemporary economists, Veblen opines that "a sequence — and a developmental process especially — must be apprehended in terms of a consistent propensity tending to some spiritually legitimate end. When facts and events have been reduced to these terms of fundamental truth and have been made to square with the requirements of definitive normality, the investigator rests his case. Any causal sequence which is apprehended to traverse the imputed propensity in events is a 'disturbing factor'." (p.220).

Veblen assesses the degree to which a sea-change within economics, analogous to that experienced within biology during the Darwinian revolution, might lead to an improved economic science. He characterises the move as one from a taxonomic framework, in which data is collected and arranged according to some coherent, yet essentially arbitrary, organising principles, to a genetic framework, in which data is collected and accounted for as the result of some unfolding economic life process. The advantage of the latter approach over the former, more "dismal" science lies in its ability to address the concerns of those interested in predicting, or explaining, extant economic phenomena, rather than merely arranging these phenomena within some pre-given scheme of norms, and accounting for irregularities as resulting from the presence of some "disturbing factor".

Veblen considered that, given his impression of the then current trends within economics, "it is only a question of time when that (substantially animistic) habit of mind which proceeds on the notion of definitive normality shall be displaced in the field of economic inquiry by that (substantially materialistic) habit of mind which seeks a comprehension of the facts in terms of a cumulative sequence." (p.240)

Thus, at the turn of the century, both Dewey and Veblen were in agreement as to the utility of the application of an evolutionary perspective both within biology and the humanities. However, both authors point out that the acceptance of a cumulative process of adaptive change (i.e., the adoption of an evolutionary perspective), may be accompanied by a teleological hangover in much the same way that the dogmatic flavour of Platonic forms tainted Kepler's empirical triumph.

Dewey identifies Asa Gray as one who, in attempting to reconcile natural selection with the argument from design, was tempted to propose a guiding hand within the evolutionary process itself. For Gray, rather than individual species taking care of their own through the exertion of some piloting influence, natural selection took the role of purposive designer, steering a stream of variations such that a final end (the natural order which surrounds us) would be achieved, and maintained.

Veblen identifies a similar intermediate position in which an attempt is made to reconcile evolutionary processes with the notion of a progressive movement towards, and subsequent maintenance of, some natural order.

> "There is no abrupt transition from the pre-evolutionary to the post-evolutionary standpoint. Even in those natural sciences which deal with the processes of life and the evolutionary sequence of events the concept of dispassionate cumulative causation has often and effectively been helped out by the notion that there is in all this some sort of a meliorative trend that exercises a constraining guidance over the course of cause and effects. The faith in this meliorative trend as a concept useful to the science has gradually weakened, and it has repeatedly been disavowed; but it can scarcely be said to have yet disappeared from the field" (Veblen, 1898, 223).

3.2 Meliorative Trends

Ironically, students of the Darwinian revolution have recently pointed out that the seeds of this teleological fallacy exist within Darwin's own work, and furthermore, seem positively exacerbated by the arguments presented by Darwin and associated evolutionists.

In discussing the role played by the idealised norms, and end-directed propensities within theories of economic behaviour, Veblen considers the possibility that their status is merely metaphorical; that such terms are glosses on underlying, non-teleological processes, which, at root, are those responsible for the outward manifestations of orderly economic behaviour (e.g., might Adam Smith's "great wheel of circulation" be best considered as a figurative reference to more proximal financial processes?).

His conclusion, that such "metaphors are effective, both in their homiletical use and as a labour saving device — more effective than their user designs them to be" (p.225–226), is echoed by Beer (1983). In her appraisal of the purposive rhetorical devices employed within Darwin's work, she extends Veblen's aside, suggesting that "the unused, or uncontrolled, elements in metaphors such as 'the struggle for existence' take on a life of their own. They surpass their status in the text and generate further ideas and ideologies. They include 'more than the maker of them knew'."

In developing this argument, Beer is amongst a group of contemporary historians, and philosophers of science, who reject the notion that the success of Darwin's theory can be entirely accounted for on the basis of its scientific merit (e.g., Morton, 1984; Young, 1985; Ruse, 1988; Bowler, 1989).

In attempting to unravel the tendency within lay folk and academics alike to read a progressive disposition into the evolutionary process, such researchers often fret over the confusion created by several of the terms and phrases most central to Darwinian theory. They worry over the threat of tautology presented by Spencer's 'survival of the fittest', the purposive implications of Darwin's 'natural selection' and 'struggle for existence', the teleological uses to which the word 'evolution' was put prior to its adoption by adaptationists[1], etc.

---

[1] For example, Beer (1983, 15) draws attention to the fact that "during the eighteenth century, when the word [evolution] was used, it meant the stages through which a living being passes in the course of its development from egg to adult", i.e., a phenomenon loaded with direction and purpose.

In addition, such authors concern themselves with attempting to identify Darwin's own position with respect to the possibility of progressive evolution, weighing his explicit rejection of the notion, which may be found repeatedly within his private correspondence and professional writings, against the famous caveat, withheld until the last passage of the *Origin of Species*, in which he paints, with one sentence, a rosy picture of natural selection as a benevolent natural process bringing about increasingly harmonious accord, contrasted with the rather less bountiful laws of physics.

> "There is a grandeur in this view of life, . . . that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved", (Darwin, 1859).

Although Darwin is typically found to be on the side of the great and the good with respect to his private notion of non-progressive evolution, he is accused of muddying the waters with his imagery, and confusing the issue with his turn of phrase. Morton (1984) neatly summarises this position using a distinction between the 'Darwinian' and the merely 'Darwinistic', which he attributes to Morse Peckham. Darwinistic arguments, whilst purporting to be under-written by Darwinian evolution, retain a teleological account of evolutionary progress which is incommensurable with Darwin's true theory. "In such a terminology" Morton claims, "a good proportion of Darwin's own work . . . would have to be condemned as Darwinistic rather than Darwinian" (p.62).

A general conclusion which has been reached within this area of enquiry is that Darwin's account of evolution admits of many interpretations, of which a non-progressive picture of adaptive change is only one. For Morton (1984), "the question at once presents itself whether any account of nature rooted in empirical observation could really permit this range of interpretation and retain its integrity" (p. 62). However, some researchers within this field identify this very multiplicity of meaning within Darwinian (and Darwinistic) writing with its subsequent success (Beer, 1983; Young, 1985).

For Beer (1983, 114), "Darwin's theories . . . contained *within them* opposing stories . . . His theory could be extrapolated to suggest a random disordered play of forces, or it could be made to yield the assurance of irreversible upward growth (his own image of the *tree* emphasised verticality) . . . The multiplicity of stories implicit in evolution was *in itself* an element in its power over cultural imagination: what mattered was not only the specific stories it told, but the fact that it told many and diverse ones."

Within biology itself, this plurality of meaning within Darwinian theory has manifested itself as a debate concerning the possibility of placing an arrow on evolutionary change (Maynard Smith, 1970; Nitecki, 1988). Although the consensus of opinion within the debate is currently that no such temporal polarity can be placed on evolutionary change, within contemporary biology as a whole there is much allusion, or explicit reference, to an ultimate directionality within evolutionary adaptation

For example, Dawkins and Krebs (1979) conclude their discussion of "escalating co-evolutionary arms races" with the speculative suggestion that such arms races may account for the appearance of a trend towards ever more complex organisms within nature. This discussion proceeds with a dismissal of the available empirical evidence (Gould, 1977), and theoretical argument (Maynard Smith, 1970), both of which resist the "commonsense feeling that there is temporal polarity in evolution” (Dawkins & Krebs, 1979, 508), before asserting that, due to the fact that organisms adapt to their biotic environment in addition to any abiotic environment, evolutionary adaptation will be progressive, at least in the short term.

In fact the notion that arms races necessarily escalate is suspect, since co-evolution is as likely to lead to cyclic trends with no progressive element, or trends in which one side achieves ultimate dominance over the other, or trends in which both sides reach some compromise. Additionally, any period of directional change which might occur during a co-evolutionary episode is not guaranteed to be ‘progressive’ by our standards, but could conceivably involve an increasing *reduction* in complexity on the part of both co-evolving populations, as, for example, each is selected for on the basis of its ability to be more stupid, or smaller, or less innervated or encephalised, than the other. Furthermore the assumption that there is a trend of increasing complexity evidenced by the fossil record, which must be accounted for by evolutionary theory, is also regarded as suspect (Gould, 1989).

The ability of evolutionary theory to account for a multitude of possible adaptive histories is its strength. In a sense this strength results in the inability to directly test evolutionary theory with observations from nature. The fact that a certain creature exists, or that a certain creature does not exist, or that a certain creature has never existed can all be accommodated by the broad sweep of evolutionary theory, although such facts may have implications for specific accounts of particular evolutionary phenomena.

This strength also results in an inability to place an arrow on evolutionary change. Raup (1988) encapsulates this argument when he considers whether the fossil record might contain clues as to evolution’s ultimate directionality. After reviewing evolutionary accounts of the fossil record which invoke evolutionary phenomena as diverse as the ‘ever-branching tree of life’, the ‘winnowing of less well-adapted forms’, the ‘rapid proliferation of new niches following extinction events’, and the ‘parallel convergent trends experienced by different lineages as they settle upon analogous solutions to some evolutionary problem’, Raup asks “If the historical record were inverted, would we recognize it as being in conflict with theory?” (p.301).

With such a panoply of selective metaphors from which to choose, each legitimised by evolutionary theory, each suggestive of differing trajectories through evolutionary space, Raup concludes that almost *any* proposed history of adaptations could be accounted for legally. Whether a hypothetical evolutionary history primarily involves periods of diversification, or periods of convergence, or is characterised by any intervening ratio of the two, metaphors for adaptation (and the associated underlying evolutionary processes which substantiate these metaphors) exist which are able to account for the fossil record that such a history would produce.

From this conclusion it appears clear that evolutionary *theory* does not *prescribe* any particular directionality upon adaptation. However, as the discussion presented by Dawkins and Krebs (1979) reveals, biological observation does seem to suggest such a directionality. However, Gould (1996) has recently presented a compelling argument suggesting that the appearance of the evolutionary process as purposive, teleological, and progressive, may be explained through a combination of three simple factors.

First, as humans we have a human-centred perspective on nature, and therefore afford the presence or absence of human-sized, human-tempo, creatures more weight than that of lifeforms more alien to us, e.g., bacteria. Secondly, the planet's modal organism, the bacterium, has not increased in complexity since its arrival, and has been the modal organism since very near the beginning of life on earth. Although the *mean* complexity of life on earth *has* increased gradually, this is merely a statistical artefact resulting from the fact that only the upper tail of the complexity distribution is free to expand. Thirdly, humanity is arrogant, self-serving, and, to paraphrase Huxley's reply to Archbishop Wilberforce, would rather remain ignorant than stomach unpalatable truths.

Of these, the statistical factor is perhaps the most substantive. If evolution is inherently progressive, or even progressive 'on the average', Gould asks, why are increases in complexity so infrequent? (See Maynard Smith & Szathmáry, 1995, for an account of the major increases in complexity which have ocurred during the course of terrestrial evolution.)

3.3 Evolutionary Perspectives Entertained Within the Non-Evolutionary Sciences

To summarise the position outlined above, evolutionary thought offers a powerful new perspective from which to assess many aspects of the natural world. The source of this power lies in the presentation of a mechanism which, although itself directionless, is capable of shaping systems over time such that they give the appearance of deliberative design. However, the power of the evolutionary perspective is offset by a tendency to view evolution as involving some meliorative trend, some beneficial tendency towards promoting what to our eyes is an appropriate state of affairs. These two themes account for the prevalence of evolutionary approaches within non-evolutionary sciences, such as linguistics, economics, psychology, anthropology, and even cosmology.

The following sections provide brief reviews of work within non-evolutionary sciences which nevertheless possess an evolutionary flavour. Subsequently two sub-classes of this work will be discussed — the application of biological evolution to phenomena typically regarded as external to biology (e.g., evolutionary explanations for phobias, or other psychological traits), and the invocation of non-biological evolutionary explanations for adaptive change within systems of agents (e.g., explaining the character of the universe through invoking some cosmological analogue of natural selection).

Once this discussion has been presented, an attempt will be made to provide some criteria with which to assess the validity of taking the evolutionary perspective on some natural system, and relatedly the validity of applying evolutionary modelling techniques to test theories concerning such a natural system.

3.3.1 Evolutionary Economics

Despite Veblen's (1898) enthusiastic prediction, at the end of the last century, that the inception of an evolutionary economics was imminent, the rate at which evolutionary ideas have been adopted by economists has been relatively slow, especially when compared to the extent that economic notions have influenced evolutionary biology (e.g., Maynard Smith, 1982).

In place of evolutionary accounts of economic dynamics, dissipative models of economic systems have tended to be the vogue. A stereotypical example of such a model is the Philips Machine. This machine was a physical implementation of a model of a national economy realised as a hydrological system of flasks and tubes, pumps and valves. Within such a model, fluid represented currency, which could flow between various reservoirs which represented public spending, national reserves, etc. Increasing interest rates might be conceived of as increasing the pressure within a certain vessel which might have ramifications throughout the system. However, the dissipative nature of the model ensured that the system tended to restabilise after such a perturbation. Only if radical interference pushed the system far from equilibrium would it not return to a stable state.

Thus, the Philips Machine includes no notion of continued economic change. The economy is conceived of as a fixed system resisting perturbations, much as Aristotle conceived of species. Enquiring of the Philips Machine how industrial innovation or positive feedback within futures markets might influence the national economy is thus a fruitless activity. The model has no room for processes other than those which are homeostatic.

However, recently a journal devoted to evolutionary economics has been established (e.g., Dosi & Nelson, 1994), and several texts stressing the utility of an evolutionary approach to economic phenomena have been published (e.g., Nelson & Winter, 1982; Foster, 1989; Epstein & Axtell, 1996). Specifically, notions of positive feedback reminiscent of Fisher's runaway sexual selection mechanism, and of the role of innovation, and the concept of economic niches have played a large part in this burgeoning literature.

Accompanying this growing interest in evolutionary accounts of economic change, contemporary authors have stressed the persistent spectre of the teleological fallacy within this work (e.g., Hodgson, 1991), vindicating Veblen's contention that purposive accounts of economic change were endemic within the economics community.

In addition to economic work which employs evolutionary notions as a useful conceptual framework within which to explore orthodox economic models (e.g., Nelson & Winter, 1974, compare neoclassical theories of economic growth with a notion of economic development couched in evolutionary terms), several studies of adaptation within economic systems have involved the construction of evolutionary simulation models (e.g., Arthur, 1993; Gode & Sunder, 1993; Dawid, 1996).

In particular, Binmore (1987, 1988, 1994) has pioneered an evolutionary framework within which to model the adaptation of economic behaviour within various games, markets, and auctions. His models are presented within the game-theoretic paradigm, and often involve the implementation of simulations in which "replicator dynamics" govern the evolution of the system through economic strategy spaces. These spaces typically feature multiple Nash equilibria. In order to generate predictions concerning equilibrium selection, i.e., to determine the likely eventual outcome of a period of evolution, Binmore models the behaviour of these systems over various time scales, and subject to various mutation regimes, in addition to deriving the number and character of the available equilibria analytically where possible.

3.3.2 Evolutionary Linguistics

A recent interest in evolutionary accounts of language origins, and language development may be identified within the linguistics, psycholinguistics, and neurolinguistics literature (Bickerton, 1990; Pinker, 1994; Jackendoff, 1993). This interest is being fuelled, in part, by genetic studies demonstrating that certain language disorders may be influenced directly by certain genes (Gopnik, 1994). Distinct from these studies concerning the evolution of language are models of language change.

For example, Kirby and Hurford (1997) demonstrate, using an evolutionary simulation model of the interaction between language learning and language evolution inspired by Hinton and Nowlan's (1987) exploration of the Baldwin effect, that the languages spoken by communities in which language structure is innately specified will not change over time.

Their model is motivated by the observation that the languages which are spoken across the globe today are not a representative sample of logically possible languages. In contrast, extant languages tend to be of a certain kind. It has been proposed that the distribution of spoken languages might be explained through some appeal to the increased efficiency of these languages over other logically possible languages. For such differences in efficiency to impact upon the distribution of language structure evidenced by contemporary living languages, a selective mechanism capable of favouring the existence of more effective language structures must be demonstrated. That is, some selective process must have ensured that certain languages proliferated at the expense of less well-adapted alternatives.

With the formulation of this theory, we see the beginnings of an evolutionary account of language change, not at the resolution of phylogenetic change, nor ontogenetic change, but somewhere in between, at a level which has been dubbed glossogenesis (Kirby & Hurford, 1997). The model provided by Kirby and Hurford (1997) explores the interaction between adaptation at each of these levels, in an attempt to derive some conditions which must be met in order that adaptive glossogenetic change might occur.

Within the model, language users are converged upon a certain language structure, and their ability to speak and understand languages of this structure has implications for their fitness. Under conditions in which this language structure is innately specified, mutant language users will suffer a loss of fitness, since they cannot speak or understand the language of their contemporaries. Such mutants will experience this loss of fitness despite any functional superiority that their mutant language might enjoy in terms of, for example, its reduced demands on short term memory.

However, allowing language parameters to be learnt during an individual language user's development introduces the ability of a language user to be born with a different innate language structure, yet still to be able to converse with other members of its language-wielding community. Within such scenarios, a mutant language user with a partially genetically specified language structure, which enjoys a functional superiority over the language structure of its conspecifics, might enjoy an *increase* in fitness resulting in the transmission of its language parameters at the expense of the less well-adapted language with which it is competing.

For example, a mutant language-user for whom a parameter that its conspecifics must learn is genetically specified may enjoy an increased ability to learn the language spoken within its community. The fitness increase enjoyed by such a mutant results in the proliferation of the gene specifying this parameter, at the expense of genes which specify that this parameter must be learned. Over evolutionary time canalization, or genetic assimilation, will result in populations fixated at the locus of this gene such that the parameter coded for at this locus is specified to be that of the mutant.

In contrast, mutant individuals whose grammar, subsequent to some critical learning period, generates sentences that are easier to parse than those generated by its conspecifics will also enjoy an associated increase in fitness. Furthermore, subsequent generations of language learners will be exposed to these more easily parsed sentences, and thus may learn more efficient grammars. This adaptive language change is what Kirby and Hurford (1997) refer to as “glossogenetic change”.

The authors’ finding is clearly related to the Baldwin effect in that their simulation model explores the interaction between language learning and language evolution. Crudely, their conclusion can be cast as a claim that without a learnt element to language no mutant language user can make themselves understood, and thus language cannot change adaptively. With a learnt element to language structure (i.e., grammar), the twin effects of genetic assimilation (i.e., the Baldwin effect) and glossogenetic change, are opposed.

A similar, but more complex model, involving closer attention to linguistic theory has been presented by Briscoe (1997). Such models of adaptive language change are clearly evolutionary simulation models despite their focus on linguistic selection.

3.3.3 Evolutionary Psychology and Social Anthropology

Psychology and evolution have enjoyed a complex and inter-related history which I do not claim to be any particular authority on. However, an evolutionary perspective on human behaviour seems to be the logical consequence of (i) the application of evolutionary thought to the animal kingdom, coupled with (ii) the fact that man is a member of the animal kingdom. Despite this apparent solid basis for inquiry, the pursuit of evolutionary accounts of human behaviour has been surprisingly unstable.

Darwin famously included very little mention of humanity in his *Origin of Species*, reserving his account of human evolution for a subsequent book (Darwin, 1871). Darwin’s protegé, Romanes (1883, 1888) explored the notion of psychological continuity between man and beast in his *Animal Intelligence*, and *Mental Evolution in Man*, an exploration which is being pursued by, amongst others, Griffin (1976, 1984, 1992) under the auspices of Cognitive Ethology (see Ristau, 1991, for a review of this field).

More direct application of evolutionary theory to human behaviour has been given many names including comparative psychology, sociobiology (Wilson, 1975), and more recently evolutionary anthropology (Boyd & Richerson, 1985), and evolutionary psychology (Barkow, Cosmides, & Tooby, 1992). Research within these fields attempts to provide adaptive explanations for the character of human culture and the human psyche.

Unfortunately, such accounts are often characterised by their opponents as ‘just-so-stories’ (for a discussion of sociobiology in this light, see Ruse, 1985). We have already seen how different selective mechanisms may be invoked to provide diverse accounts of evolutionary trajectories (see section 3.2). The root of the repeated objections to evolutionary accounts of the origin and character of human behaviour seem to be fuelled by the apparent arbitrariness of these accounts, and their tendency to support rather crude pictures of human societal order. Such objections persist despite the most recent evolutionary psychology distancing itself from what was perceived as the crude genetic reductionism of sociobiology.

Researchers within these disciplines have scant means of testing their verbal arguments, and little inclination to do so. As a result there have been only a few attempts to model the evolution of psychological systems (although see Miller & Todd, 1990; Todd & Miller, 1991a, 1991b, for evolutionary simulation models of various learning processes).

The relationship between developmental psychology and embryology, on the one hand, and evolutionary biology on the other is perhaps more intimate and complex than any other psychology/biology relationship (e.g., Butterworth, Rutkowska, & Scaife, 1985).[2] Costall (1985) discusses the widespread use of developmental evidence to fallaciously support phylogenetic gradualism in the wake of the publication of the Darwin's works on evolution (what he terms a "lazy man's palaeontology", p.35), and the subsequent rejection of Haeckel's (1874) overly simple notion of ontogeny recapitulating phylogeny in favour of developmental theories with no evolutionary element. Piagetian genetic epistemology (Piaget, 1972) seemed to revitalise a relationship between phylogenetic and ontogenetic change, but, Costall maintains, involved confusions as to the relationship between organism and environment over phylogenetic and ontogenetic time.

Current progress in evolutionary theory has enabled developmental psychologists to move forward from Haeckel's (1874) simplistic recapitulation theory. In its place a notion of ontogenetic development is being constructed which, in addition to being constrained by both genetic influences and environmental structures, and in contrast to evolutionary development, is conceived of as a directed process, in the sense that there is an adult organism which *should* arise from an embryo, not merely one that *will* arise.

This dynamic, adaptationist account of development has yet to be adequately captured by theory, but several initial moves in this direction have been made (e.g., Thelen & Smith, 1994).

Dawkins (1976) has suggested that ideas may themselves evolve, competing for the limited resources of (human) memory and attention. Although the details of mimetic evolution are sketchy, the notion has had considerable appeal, and recently attempts have been made to analyse the development of discourses in terms of memetic competition (e.g., Best, 1997). However, until the details of memetic reproduction, and heritability, selection, etc., have been determined such models appear to be at best premature.

These particular intersections of evolutionary theory with social and psychological sciences may be regarded as analogous to the intersection of evolutionary theory and linguistic development described above, in that, rather than involving an interest in the biological evolution of human linguistic or developmental features, researchers within these intersections attempt to apply evolutionary, adaptive, and dynamical notions to theories of change within a non-phylogenetic timescale.

3.3.4 Evolutionary Cosmology

I have included a discussion of "evolutionary cosmology" as an example of an area of scientific enquiry within which the adoption of an evolutionary perspective seems less well-advised. This discussion draws primarily upon a short commentary recently published by Maynard Smith and Szathmáry (1996).

Several physicists have proposed that the structure of the universe we inhabit seems almost miraculously suited to our needs as human beings. This has lead some to propose that our existence now has somehow influenced the manner in which the universe 'evolved' way back when (e.g.,

---

[2] As mentioned in a previous footnote, the term evolution was initially used to refer to ontogenetic development.

Barrow & Tipler, 1986; Smolin, 1997). The disposition of such authors to account for what they consider to be a comfortable natural order through some appeal to the necessity of this state of affairs smacks of a teleological foundation, and, to their credit, authors such as Barrow and Tipler (1986) pay some considerable attention to the role of teleology within their ‘anthropic cosmological principle’.

One rendering of the notion that human life within the cosmos is a likely, or even necessary, product of the way physics works has been presented by Smolin (1997). Smolin proposes that our universe is the product of cosmological evolutionary processes which select for certain kinds of physical laws, laws which just happen to be able to support human life. However, Maynard Smith and Szathmáry (1996) have pointed out that one of the ingredients required by an evolutionary account is missing from this notion of cosmological evolution.

In order that some form of selection over possible worlds is possible, these possible worlds must compete for some limited resource. Although this stipulation does not require that evolving systems come into direct physical contact with one another, it does require that the increasing success of one system necessarily results in the decreasing success of others. Since multiple universes may be infinite in number, there seems no equivalent to the limited resources which terrestrial organisms compete for (although Gribbin, 1997, has suggested that these universes are “jostling for spacetime elbow room within superspace”, where superspace contains all possible causally isolated universes). Without this analogue, multiple universes are free to develop in any manner; they are free from the constraining hand of selective forces. As Maynard Smith and Szathmáry (1996) point out, although one may predict the rates of reproduction of these possible worlds, one cannot claim that particular possible worlds will exist at the expense of others.

3.3.5 Evolutionary Computation and Evolutionary Design

The application of evolutionary principles to the optimisation of man-made systems (e.g., maximising the through-put of a production plant through evolving efficient production schedules, Husbands, 1993) necessarily involves an end-directed notion of selection. Engineers implementing evolutionary search algorithms do so with a purpose. They have a goal to which they hope evolution will lead them. However, although, as discussed above, imputing an inherent progressive tendency to natural evolution is contentious, as an engineering practice, a reliance on artificial selection as an optimisation technique is perfectly valid.

Whereas natural selection does not strive toward any goal, it is possible, through judicious interference with natural evolution, to *artificially* steer evolutionary change in a certain direction. Artificial selection as practiced by animal breeders is exactly this judicious interference. Since the breeder is able to steer the evolutionary process through imposing a selective breeding regime on her stock, goals (such as a strain of animals with a particular milk yield) may be achieved over evolutionary time. Indeed, the example of directed change resulting from artificial selection led Darwin to propose the term natural selection as an analogous term with which to describe the undirected equivalent which takes place in the wild.

Many researchers implementing evolutionary algorithms as optimisation processes appreciate that the closest analogy to their work is to be found within artificial selection paradigms, rather than evolution in the raw (e.g., Husbands, 1992). However, within the evolutionary computation field there is a tendency to attribute the power of evolutionary 'design' (which engineers wish to harness, and turn to their own ends) to various features of natural selection which are absent from artificial selection. This tendency is probably defensible. Natural selection has demonstrated its ability to evolve man from monad, whereas artificial selection upon a particular breed seems limited to varying but a few crude characteristics, and even then only within certain bounds.

The limits of artificial selection are accepted by some 'evolutionary engineers'. For example, Harvey et al. (1993), in a paper promoting the use of artificial evolution as a technique capable of designing controllers for autonomous robots, maintain that "when using the evolution of animals as a source of ideas for the evolution of [robot controllers, evolutionary algorithms] should be used as a method for searching the space of possible adaptations of an existing [robot controller], not as a search through the complete space of [robot controllers]" (p.365). This engineering approach based on animal breeding should perhaps be termed artificial husbandry in order to distinguish it from artificial selection as it is practiced by animal breeders (i.e., natural husbandry).

Although features of evolution in nature such as co-evolution, speciation, etc., are indubitably responsible for the extended ability of natural selection, with this extended ability comes a feature of natural selection which should be anathema to 'evolutionary engineers' — its directionlessness.

Although many papers on co-evolutionary design have been published within the evolutionary computation field (e.g., Hillis, 1990; Reynolds, 1994; Paredis, 1995; Sims, 1995; Cliff & Miller, 1996; Pollack, Blair, & Land, 1996), as yet no technique for circumventing the purposelessness of co-evolutionary change has been developed. Authors have developed techniques for judging the presence, rate, and extent of evolutionary progress within co-evolutionary design regimes (e.g., Cliff & Miller, 1995), and have speculated as to the character of problems which will prove amenable to co-evolutionary styles of automatic design (e.g., Pollack et al., 1996), but these considerations are rare. For the most part, the literature may be characterised as resulting from an optimistic belief in the ability of co-evolution to produce complex, robust solutions to difficult design problems, coupled with a grudging acceptance that co-evolutionary design may need special 'kludges' designed to ensure continual progress in some preferred direction.[3]

Such kludges have typically been arrived at in an unprincipled and atheoretic manner. It seems reasonable to suggest that a more profitable approach might be to derive techniques for encouraging directional evolutionary change from theory within evolutionary biology. Such techniques may not *guarantee* continual progressive evolution toward some specified goal, but may increase the likelihood of directed evolutionary change, at least for limited periods. This topic is explored further in chapter 4.

3.4 Evolutionary Models of Non-Evolutionary Phenomena

The research sketched above may be divided into two categories. First, some work applies the theory of biological evolution to phenomena which, although natural and at root biological, are not typically addressed by biologists. Such research attempts to account for the origin and subsequent development of such phenomena from a biological evolutionary perspective. Examples of this

---

[3] For example, both Rosin (1997) and Cliff and Miller (1996) assert that co-evolutionary design is assisted if current solutions are assessed with respect to both their *current* co-evolutionary partners and the best of their *ancestral* co-evolutionary partners, i.e., a population of ancestral elites which is maintained artificially over evolutionary time.

class of research include discussions of the origins, and evolution, of language, and psychological phenomena.

Such research is problematic in that verbal arguments concerning the evolution of complex traits, sometimes specific to humans, are hard to formalise as clear models.

The second category of research involving the adoption of an evolutionary perspective involves attempting to explain the dynamic characteristics of systems which may be non-biological through an appeal to some adaptive mechanism which, although not biological evolution, bears some resemblance to it. Such work includes models of glossogenetic change, memetic evolution, cultural evolution, and potentially cosmological evolution.

Before such research can be carried out successfully, the extent of the analogy between biological evolution and the 'evolutionary' process posited as an explanation of dynamic change within the target system must be described. This description must address many issues, including the nature of heritability within the target system, a characterisation of the units of selection within the system, the resources for which the agents within the system are competing, etc.

For example, might memetic evolution involve Lamarckian inheritance? Should economic learning be characterised as evolutionary change driven by competition for goods, money, or 'utility'? Might glossogenetic change involve both genes and memes? When considering evolutionary cosmology, might the 'genes' which code for the laws of one universe behave analogously to biological genes when that universe spawns another (i.e., particulate inheritance), or might such 'genes' mix together in some continuous fashion (i.e., blending inheritance)?

Differing answers to questions such as these will have profound effects on the evolutionary dynamics of the systems at issue. For example, if mutant memes are distributed around their parents in a directed manner, i.e., if new memes tended to be different from their parents in a direction which tends to increase their fitness, the resulting evolution would be radically different to that of biological organisms.

As Maynard Smith points out in his *Evolution and the Theory of Games*, game theoretic modelling techniques are capable of capturing the dynamics of cultural evolution and developmental evolution, in addition to biological evolution. Similarly, evolutionary simulation models are suitable tools for approaching both of the categories of research described above. However, before such modelling can be achieved profitably, the mechanisms governing evolutionary change must be clearly specified.

The use of evolutionary inspiration in the creation of automatic design processes fits neither of the two categories outlined above since both are scientific enterprises, whilst artificial evolutionary design is an engineering practice. As such, researchers interested in evolutionary design are free to explore any possible answer to the kind of questions posed above. The only constraint on these decisions is that they result in effective evolutionary design. For example, researchers within this area have explored Lamarckian inheritance (Sasaki & Tokoro, 1997), sexual reproduction involving more than two partners (Coker & Winter, 1997), sexual selection (Miller, 1994), and the Baldwin effect (Mayley, 1997), with a view to increasing the efficiency of evolutionary design.

3.5 The Evolutionary Perspective and Evolutionary Simulation Modelling

In this chapter, the evolutionary perspective was first described in terms of a rejection of animist Greek philosophy, a philosophy steeped in teleology. Subsequently, the degree to which Darwinian thought actually constitutes emancipation from this teleological position was discussed, before a review of evolutionary perspectives in the non-biological sciences was presented.

It should by now be clear that in order for the evolutionary perspective to be validly adopted toward some system, it must be the case that the system comprises agents competing for some resource. Furthermore, such agents must reproduce and their reproduction must include some element of heredity. Such a system will evolve over time and exhibit a propensity to adapt to changing circumstances in a manner which tends to increase the adaptive fit of the agents to their environment. A system to which one may validly adopt an evolutionary perspective is one which may profitably be explored using evolutionary simulation modelling.

Chapter 4

Prospects for Artificial Co-Evolutionary Design

Genetic Algorithms (GAs) typically work on static fitness landscapes. In contrast, natural evolution works on fitness landscapes that change over evolutionary time as a result of (amongst other things) co-evolution. The attractions of co-evolutionary design techniques are discussed, and attempts to utilise co-evolution in the use of GAs as design tools are reviewed, before the implications of natural predator-prey co-evolution are considered. Utilising strict definitions of *true* and *diffuse* co-evolution provided by Janzen (1980), a distinction is drawn between two styles of evolutionary niche, Predator and Parasite. The former niche is robust with respect to environmental change and features systems that have had to solve evolutionary problems in ways that reveal general purpose design principles, whilst the nature of the latter is such that, despite being fragile and unsatisfactory in these respects, it is nevertheless evolutionarily successful. It is contested that if co-evolutionary design is to provide systems that solve problems in ways that reveal general purpose design principles, i.e., to provide robust styles of solution, true co-evolution must be abandoned in favour of diffuse co-evolutionary design regimes.

4.1 Introduction

This chapter will act as an interlude, dividing the largely philosophical material which precedes it from the largely theoretical material which follows it. Within this chapter the engineering practice of artificial co-evolutionary design will be described and assessed from the perspective developed in the last chapter. This discussion will also serve to introduce the notion of a Genetic Algorithm (GA), a computational algorithm which captures the logical structure of evolution. Such algorithms will feature heavily in the evolutionary simulation models developed in chapters 6 and 7.

The attraction of artificial evolutionary design techniques lies in the undisputed majesty of naturally evolved systems. That the organisms which populate the world – complex, economical, and robust solutions to evolutionary problems – were generated through a process of replication with variation is a supremely encouraging discovery. The development of man-made design processes fashioned in evolution's likeness offers us the chance to harness and direct some of the creative power of evolution.

However, the most representative products of evolution are not necessarily the organisms that catch the evolutionary roboticist's eye. For example, evolution does not of necessity strive for complexity. This is attested to by the multitude of simple organisms prevalent on this planet. Similarly, although economy is the norm amongst evolved creatures, there exist selective pressures which are capable of producing energetic, opulent and costly extravagance (Fisher, 1930; Zahavi, 1975, 1977a). Further, although the vast majority of the solutions to life's trials arrived at via the evolutionary process *are* robust systems towards which, as a designer, one might direct one's admiration, there is evidence to suggest that evolution is equally capable of generating systems which should fail to satisfy the evolutionary roboticist's project.

Consider the problem encountered by Clever Hans, the counting horse. Hans was required to tap his foot a number of times to indicate the answers to simple sums. Hans arrived at a solution to this problem that I will term 'fragile'. Rather than actually performing the calculation, Hans merely tapped his foot until his trainer gave a slight, unintentional signal. As Hans' trainer always gave this signal when Hans had tapped the correct number of times, Hans always got the answer right. This solution is fragile because trivial changes in the problem which should not affect Hans' performance, such as removing the trainer from Hans' line of sight, will destroy Hans' 'mathematical' ability. Note that Hans' solution is still robust under a fairly wide range of circumstances including changes in the weather, the clothing of his trainer, etc. Fragility is thus context-dependent, and a matter of degree.

Evolution appears to have weeded out fragile solutions, having demanded of its creations that they satisfy reproductive demands in a wide variety of hostile, unpredictable situations, and indeed this is, on the whole, the case. However, such robustness is not a product of the *evolutionary process*, so much as the *selective pressure* under which most of natural evolution's subjects have toiled. In exceptional cases, evolution can produce systems as fragile as Hans' claims to numeracy.

If such systems exist in nature (and I will attempt to show that they do), how are we to avoid generating them in our attempts to evolve systems which are solutions to design problems? How can an evolutionary roboticist guarantee robust solutions? In answering these questions it will be useful to consider evolutionary biology's notion of co-evolution and how artificial evolution measures up to it.

4.2 Co-evolutionary Design

Genetic Algorithms (GAs) are a design/optimisation technique inspired by natural evolution (Holland, 1975; Goldberg, 1989). The bare essentials of evolutionary theory (selection, reproduction, variation, fitness, etc.) are idealised and applied to artificial genetic material in an attempt to evolve solutions to problems.

A genetic algorithm works on a group of potential solutions to a problem, termed a population. Each solution takes the form of a string (chromosome) of letters (genes) from an alphabet (typically consisting of the binary digits $\{0,1\}$). The algorithm first assesses these 'genotypes' allocating a fitness value to each, dependent on the degree to which its associated 'phenotype' solves the problem (e.g., how well the parameters specified by a chromosome optimize a function, or characterise a useful bridge, or pack a lorry, etc.).

Once this has been carried out, a new population of genotypes can be created by breeding the current population. Parent chromosomes are chosen with some bias towards those that are most fit. Reproduction may be asexual, in which case a parent genotype is merely copied to its offspring, or sexual, in which case each parent contributes some genetic material to their offspring. This is typically achieved through some form of cross-over operator which takes two chromosomes and produces two more by swapping a randomly chosen portion of the genome from each parent. During reproduction there is a small chance that mutation may occur resulting in a less than perfect copy of genetic material. The role of mutation in the reproduction process is to introduce some variation within the population. The cycle then repeats as the offspring generation is assessed, and bred, etc.

Genetic algorithms, like other optimisation techniques, are often thought of as traversing landscapes. A potential solution is represented as a point in such a landscape, the height of which corresponds to its ‘fitness’ – the extent to which it solves the problem. Fitness can either increase with height or increase with depth. For the remainder of this chapter I will adopt the former convention.

Genetic algorithms typically work with a population of solutions scattered across the fitness landscape which gradually converge on one of the fitness maxima. The fitness landscape is fixed by the designer of the GA when she decides how she will assess the potential solutions in the population. For example, if evolving a bridge design, the GA designer may specify fitness as being a function of some measure of safety, a measure of traffic capacity, and a measure of cost. As such a function does not change throughout the evolution process, the fitness landscape can be regarded as static, i.e., a potential solution with fitness x will always have fitness x independent of its peers or the passing of time. Indeed almost all optimisation techniques traverse static fitness landscapes.

A fundamental problem for the designer of genetic algorithms is specifying the problem that is to be solved, in terms of a fitness function, in a manner that allows incremental steps towards a solution to be rewarded. If the fitness function, a metric which is used to assess the degree to which a solution solves the problem, is not so constructed, the genetic algorithm may spend periods of time on plateaux with no method of discriminating between competing solutions, or favour solutions which are evolutionary blind alleys leading to local maxima from which it is hard to escape, or suffer from excessively ‘noisy’ fitness landscapes which are prohibitive of incremental progress.

In contrast to these static landscapes, natural evolution works on a dynamic fitness landscape (Bullock, 1995b). Over evolutionary time the fitness of a phenotype (solution) may change radically. What was a winning strategy (e.g., eating flora of type A) becomes ‘out of date’ as conspecifics, predators, resources, etc., change through their own evolution. The resulting *co-evolution*, the evolution of systems in response to each other, can be thought of as ensuring that an organism’s evolutionary goal-posts (maxima in the fitness landscape) move. Rather than working towards the solution of some fixed problem, organisms are constantly adapting (over evolutionary time) to each other, their surroundings, etc., which are themselves adapting in response.

The possibility of artificial co-evolutionary design techniques offers the opportunity to circumvent the problems associated with static fitness landscapes by automatically moving the GA’s evolutionary ‘goal-posts’, gradually changing the problem as the population moves over a dynamically changing fitness landscape. What are the prospects for such an automatic co-evolutionary approach? Initial work in this area is thin on the ground, but can be divided into two camps.

First, artificial co-evolutionary paradigms concerned with modelling natural phenomena such as communication include seminal work by Werner and Dyer (1991), in which male and female animats co-evolved signal production and signal interpretation protocols in a simple animat-world, allowing immobile females to attract blind males. Although research paradigms involving such idealised worlds are fraught with difficulties (Cliff & Bullock, 1993), studies such as this laid the groundwork for subsequent, more theoretical, co-evolutionary studies of parental imprinting (Todd & Miller, 1993), aggressive signalling (de Bourcier & Wheeler, 1994), predator-prey co-evolution (Miller & Cliff, 1994), sexual selection (Miller, 1994), and co-evolutionary analysis tools (Cliff & Miller, 1995).

Although such work is theoretical in flavour, as the nature of artificial co-evolution is uncovered, and the intricacies of natural co-evolution are appreciated, implications of a more practical nature, particularly in the field of evolutionary robotics, seem inevitable. However, it is to the second group of research efforts that this chapter is primarily addressed.

This second embryonic research programme is concerned with utilising artificial, co-evolutionary paradigms in the design of useful systems. In contrast to the theoretical drive of the co-evolutionary simulations mentioned above, these research projects are practical, pragmatic attempts to engineer solutions to real-world problems. In the terminology introduced in chapter 2, the distinction being made here is one of Model class vs. Working class artificial life research. Initial work in this area includes research by Hillis (1990) and Robbins (1994), in which parasites are used to increase the performance of artificial agents, Husbands (1993), in which the co-evolution of shop-floor schedules was explored, and (Pollack et al., 1996) in which backgammon players were designed through a co-evolutionary process. Such work, however, is in its infancy.

In addition, the incremental approach of Harvey et al. (1994) can be seen as an attempt to use co-evolution in the design of autonomous agents. The agents involved initially face a simple sensory-motor problem, which is incrementally made more difficult in an effort to coax complex behaviour from systems which could not be evolved from scratch. Such scaffolding techniques are reminiscent of the parent-child interactions that facilitate infant development (Rutkowska, 1994).

However, the hand-cranked nature of such scaffolding requires the presence of a human designer ‘in the loop’ and, potentially, the tasks of specifying the incremental goals that allow evolution to reach solutions to complex problems could itself become as problematic as designing the agents manually. However, a more general problem may be inherent in such co-evolutionary approaches. Before describing this problem in more detail, the natural co-evolution of predators and their prey will be considered.

4.3 Predator-Prey Co-evolution

Predator-prey evolutionary dynamics exhibit what behavioural ecologists have termed ‘evolutionary arms races’ (Dawkins & Krebs, 1979). For example, the development of higher acuity in a predator may be countered by the evolution of camouflage in a prey; teeth and claws provoke carapaces and scales; toxins demand antidotes, etc.

It is suspected that such arms races may result in highly developed behavioural skills and complicated morphology. This complexity is the result of the increasing demands placed on organisms by their environment (including their conspecifics, predators, prey, etc.). The hunting skill and speed of the peregrine falcon, for example, would not have evolved without the concurrent evolution of the perceptual capacity and escape capabilities of its prey.

Does any kind of co-adaptation qualify as co-evolution? Janzen (1980) distinguishes between true co-evolution and what he terms "diffuse" co-evolution. He defines the former as continued, *specific*, reciprocal, evolutionary change, i.e., continued evolutionary change in one specific trait of one population in response to the continued, reciprocal, evolutionary change of one specific trait possessed by another population. In contrast, diffuse co-evolution is continued, *non-specific*, reciprocal, evolutionary change, in which a trait of one population changes over evolutionary time in response to a *group* of traits possessed by another population (which may contain several species).

For example, the evolution of egg-mimicry and egg-discrimination in species of bird that respectively perpetrate and suffer the dumping of eggs in foreign nests is an example of true co-evolution in that the traits involved have evolved specifically for the purpose of brood-parasitism and defeating brood-parasitism respectively. Experimentation shows that species beyond the normal range of host-parasites show less strong egg-discrimination than those which routinely suffer egg-dumping, whilst species not subjected to such parasitism show little discriminatory ability at all (Davies, Bourke, & de L. Brooke, 1989). Furthermore, groups within a species of parasite which differ genetically may evolve mimetic egg patterns specific to different hosts.

Conversely, although the hard, ridged shells of many crustaceans are the result of the continued, evolutionary escalation of shell breaching mechanisms employed by their predators, these impregnability traits have evolved in response to a group of traits in a number of predatory species (a variety of shell piercing/crushing techniques), rather than one specific trait (e.g., pincers) and are thus examples of diffuse co-evolution (see Endler, 1991, for further examples).

In fact, under Janzen's definitions, it is hard to find *any* instances of co-adapted predator-prey traits that can be classed as the product of *true* co-evolution. For example, the speed and strength of feline predators did not evolve solely in response to the escape tactics of their prey, but under a more general, "diffuse" selective pressure that favoured faster and stronger cats for a variety of reasons. Similarly, the discriminatory abilities of predators tend to be general ones rather than the egg-specialised systems of cuckoldry sufferers.

In contrast, there is evidence to suggest that much parasite-host co-evolution is *true* co-evolution. In addition to the egg-dumping research mentioned above, ant colony parasitism by alien queens, and mammalian ectoparasites all feature co-evolved traits (Endler, 1991).

Endler explores a number of reasons to account for the differences between predator-prey and parasitic arms races. Firstly, whilst predators typically exploit many prey species, most parasites are host-specific. This ensures that specialisation by predators is difficult whilst parasites experience a "more consistent and simpler selection pressure" (p.195), allowing the genesis of specific adaptations to their host species to occur more easily. Predators practicing apostatic selection, preferentially over-exploiting the more frequent prey types, and switching between preferred prey-types as prey densities change, will also experience less selective pressure for specialised traits. As parasites kill their prey more rarely, and thus spend more evolutionary time concentrating on one species, apostatic effects are a less significant pressure.

Predators encounter generalised defences more frequently than specialised ones as they typically occur earlier in the predation sequence — the series of stages that must be passed through before a predator consumes a prey item. The predator must proceed from Encounter, through

Detection, Identification, Approach, and Subjugation to Consumption of the prey, and will be confronted with various defences along the way. For example, at the level of Encounter, hiding, alertness, and acuity are typical defences utilised by many prey types, but at the Subjugation/Consumption end of the sequence defences are typically specific to prey types (e.g., emetic unpalatability causing vomiting, spines, dispensible body parts, mobbing, etc.). This implies that adaptation to general defences may be more profitable for predators than evolving to defeat more specialised traits, but does not generalise to parasite-host scenarios.

The life/dinner principle (Dawkins & Krebs, 1979) suggests that the selective pressure on prey is much stronger than on predators. If a predator fails to secure a prey item it has lost its lunch, but if a prey item fails to avoid being secured by a predator, it has lost its life. As Dawkins and Krebs point out “A fox may reproduce after losing a race against a rabbit. No rabbit has ever reproduced after losing a race against a fox” (p.59). Again this effect is weakened in the parasite-host situation as the host is killed less frequently.

Finally, prey species benefit in their struggle with predators from their superior numbers and, typically, their shorter breeding cycle. Both these factors allow faster evolution of prey than predators. As parasite numbers may equal and in many cases exceed those of their hosts, and their breeding cycles are not necessarily longer than those of their hosts, parasite evolution may proceeded at rates equal to or greater than those of their hosts. This ensures that host adaptations may be matched blow for blow by parasites and in some cases comfortably ‘out-evolved’ by fast-breeding, numerous, parasitic organisms.

4.4 Implications

The differences between the evolutionary niches of the predators and parasites described above, and the resultant differences in both their behaviour and morphology hold a lesson for evolutionary robotics. Simplifying the findings described in the previous section, one can construct two idealised, stereotypical niche styles, Predator and Parasite. Whilst Predators cope with environmental change through recourse to flexible, general-purpose skills, Parasites have settled into niches that allow them the luxury of specialised, brittle life-styles. In the terminology of Section 1, Parasites are fragile solutions to the trials of life whilst Predatory solutions are robust [1].

What is wrong with the Parasitic lifestyle? In what sense is it brittle or fragile? It is evolutionarily successful, admirably economic, and at times breathtakingly complex. Granted, from the point of view of the behavioural ecologist the Parasite is no worse than any other style of niche. However, I would assert that evolutionary robotics is aiming for Predatory solutions to the problems it addresses rather than Parasitic ones. Parasites succeed in a manner that is inadmissible from the point of view of evolutionary robotics. As we saw in the previous section, they succeed

---

[1] It must be reiterated here that the notions involved in this discussion are not dichotomies. They are graduated. No evolved system is truly general purpose. For example, although the vision system possessed by humans is a very flexible and general mechanism, it has limits, and may throw no light on the design principles of some extra-terrestrial vision system consisting of a tri-focal, x-ray-sensitive device. Similarly, no naturally evolved system is completely parasitic. Even Clever Hans’ fragile foot-tapping displays a certain degree of flexibility. Further, judgements of fragility and robustness are context-dependent in that they are made with respect to some pre-given task or domain. If one is evolving discriminatory abilities as part of a limited-scope egg-recognition system, one may be convinced of the robustness of a quite simple, egg-specific visual system. However, if a system is evolved as part of an endeavour to design a more general visual mechanism, more of it might be required before its robustness may be asserted.

through superior reproductive advantage and relatively simple selective pressure.

These two factors allow Parasites to eschew general-purpose solutions to the problems confronting them, in favour of specialised solutions that can be changed at a rate comparable to that of their hosts' counter-solutions.

Imagine a Parasitic bridge design. It has evolved specifically to exploit the river it spans, and the traffic that crosses it. If one challenges such a Parasite with environmental change within its lifetime it will most likely fail. Such a bridge will not adapt to span a different river. "So what?", one may retort, "It was evolved to carry Traffic X over River Y." Further, if one examines such a Parasite in order to discover some general design principle, one will most likely be left dissatisfied. The characteristics of such a bridge which contribute to it being a good bridge will be indistinguishable from the characteristics which make it a good Bridge-Over-The-River-Y-With-Traffic-X.

Such solutions are analogous to that of Clever Hans. Whilst perhaps being adequate, they will not reveal general purpose principles because they do not succeed through the adoption of general-purpose principles. If artificial co-evolutionary design is to avoid generating solutions analogous to the parasites discussed above, care must be taken to avoid scenarios in which artificial agents are subjected to selective pressures which promote tightly co-adapted solutions at the expense of robustness.

Although it may be impossible to ensure that co-evolutionary design regimes always lead to robust solutions, steps can be taken to improve the frequency with which such designs are achieved. Taking a cue from natural systems, evolutionary roboticists might eschew co-evolutionary regimes in which the solution population is pitted against a single population of co-evolving adversaries. As mentioned above, such a true co-evolutionary situation tends to lead to Parasitic solutions. There is more chance of generating Predatory solutions to a problem if a variety of selective pressures are experienced by the solution population. This might be achieved if multiple populations of co-evolutionary adversary are pitted against the solution population.

A degree of support for this position is provided by Kauffman's (1993) observation of evolution over coupled fitness landscapes. In Kauffman's NKC scheme, the fitness contribution of genes within an N-gene genotype is mediated by K genes within the same genotype and C genes within the genotypes of S other populations (i.e., K controls the degree of epistasis within a genotype, whilst C controls the degree to which species are evolutionarily coupled, and S controls the extent of population coupling).

Kauffman maintains that the fitness maxima achieved by co-evolving populations are, on average, higher in the fitness landscape than those achieved by isolated populations, and that there exist optimal values of K, C, and S for any N, such that populations achieve high average fitness maxima. These values prescribe scenarios in which populations are coupled to a medium degree. Kauffman characterises this medium degree of coupling as occurring near a phase transition which divides systems which are too "static", for those which are too "chaotic". Robust solutions (i.e., optima which are high in the fitness landscape) are achieved by systems which are at "the edge of chaos".

Whatever the value of the conceptual framework within which Kauffman works, for the purposes of this chapter his research demonstrates that the degree of coupling experienced by co-evolving populations has an influence on the character of evolutionary optima achieved by a co-evolving system. In the terminology introduced by Janzen (1980), systems with low S or C are systems experiencing *true* co-evolution, whilst systems with high S or C are systems experiencing *diffuse* co-evolution.

Kauffman's results suggest that true co-evolution will lead to sub-optimal equilibria, whereas overly diffuse co-evolution may not settle on optima at all, leading to perpetual cycling of evolutionary strategies. Attempting to discover a balance between these two extremes may enable co-evolutionary design to achieve, on average, robust solutions to difficult design problems.

4.5 Conclusion

In conclusion, two broad claims have been made within this chapter. The first is that co-evolutionary design techniques are a useful way of dealing with the problems associated with static fitness landscapes. The second is that co-evolution will only produce fragile solutions unless it is of a relatively diffuse nature. I intend neither of these points to be dogmatic as they are by no means applicable to every research programme. For example, if one can construct a good static fitness landscape, one that allows a GA to move towards, and eventually reach, an optimal solution, then co-evolutionary techniques may be superfluous (although even in this case they may accelerate the optimization process). Similarly, if, for some reason, fragile solutions are acceptable, the demand for diffuse selective pressure need not be met.

Indeed, in the case of the Parasitic Bridge described in the previous section, it may well be that a fragile solution is perfectly acceptable, as the general design principles underlying bridge building are well known. No extrapolation from the Parasitic Bridge to further bridge building efforts is necessary.

However, if one's evolutionary design is in any way exploratory; if one is working in a domain which is little understood, in a domain in which early successes must justify investment in the similar procurement of later successes; if one thus wishes to make claims of solutions arrived at through evolutionary design concerning the potential for such evolutionary design, to make general claims concerning classes of solution and the underlying principles that unify them, then one had better be sure that the solutions upon which one bases such claims are not Parasitic. Because Parasitic solutions admit of no such claims. Clever Hans was not the first step on some road to a theory of animal addition.

Chapter 5

The Evolution of Natural Signalling Systems

Communication appears near ubiquitous throughout nature. Signals are used by animals (and indeed plants) for a variety of purposes, and feature a bewildering degree of variation in both form, function, and character. Signals are passed through many different media, utilising many different sensory modalities, and emanating from a variety of organs. They are exchanged by both kin and unrelated individuals, and are sent both within and between species. They may even be transmitted within an organism, between organs (e.g., messages from stomach to brain), or cells. Signals may be co-operative or exploitative, extravagant or restrained, directed or broadcast, elaborate or simple.

Within this chapter, competing theories which attempt to account for the characteristics of natural signalling will be presented, and problems in defining communication which result from these theories will be discussed.

5.1 Efficacy

Researchers concerned with natural signalling systems are faced with the problem of accounting both for the diversity of communicative behaviour, and the specific character of individual signalling systems. Although this is a daunting task, considerable headway has been made in describing both general selective pressures which might favour, for example, repetitiveness of display (e.g., Wiley's, 1983, account of redundancy within animal signals), and the specific contingencies of evolutionary history which could account for the character of individual systems (e.g., Webb's, 1994, account of cricket phonotaxis in terms of the acoustic properties of the cricket's sensory system).

Guilford and Dawkins (1991, 1993) have characterised this research effort as an attempt to account for the design of signals in terms of their *efficacy*. The efficacy of a signalling system concerns the degree to which its design enables the effective production, transport, and consumption of signals. Factors which influence efficacy will therefore often be specific to the particular characteristics of the signaller, receiver[1] and the medium through which the signal passes.

---

[1] In this chapter, and subsequent chapters, I shall use the terms receiver, and observer, to refer interchangeably to the recipient of a signal. Although many signals may not be observed, i.e, they are not visual signals, the term 'observer' serves to draw attention to the active element of signal receipt.

Under selective pressure for efficacy, signals should become adapted to be optimal channels of communication given the physical medium and environment through which they must travel, the constraints on signaller morphology and behaviour, and the sensory and psychological apparatus of the intended receiver (and possible unintended receivers). Since the latter systems are themselves subject to selective pressure, one might expect to find sensory organs which had been tuned by evolution to be especially sensitive to the signals of interest to their possessors, and production organs (e.g., the legs of the cricket, or peacock's train) which had been similarly adapted to efficiently produce appropriate signals.

The various costs associated with being involved in a signalling system, which militate against efficacy, include production costs, vulnerability costs, time costs, and the risk that signals may be degraded by environmental interference, or ignored by recipients, or mistaken for alternative signals, or overheard by eavesdroppers, etc. Selective pressure for increased signal efficacy will conflict with selective pressure for reduced signal costs. In such situations it is predicted that a balance between the two will be struck, i.e., signalling systems will tend to evolve to become maximally *efficient*.

Features of signals which have been identified as traits selected for as a result of the increase in efficacy they bring about include conspicuousness, redundancy (both spatial and sequential), stereotypy, and the presence of alerting components (Wiley, 1983).

5.2 Strategy

However, a recent change in evolutionary perspective (Hamilton, 1964), as naive group selectionist accounts of evolved traits fell into disrepute, has led to a fundamental reappraisal of the evolutionary utility of signalling in general. Prior to the reassessment of group selection during the mid-sixties, the evolutionary function of signalling behaviours could be explained in terms of benefits to the signalling community as a whole. Such communities were thought to benefit from signalling systems which efficiently co-ordinated the activity of the group. A community which effectively co-ordinated the division of resources, for example, would outperform a community which divided resources through some less efficient means, e.g., through fighting (see Sebeok, 1977; Smith, 1977, for examples of this more traditional conception of signalling).

However, a perspective on evolutionary change which afforded primacy to explanations pitched at the level of the individual, or gene, could not tolerate such accounts of signalling since the stability of such communication could be undermined by individuals able to prosper at the expense of the group, e.g., cheats, bluffers, free-riders, mimics, liars, and frauds.

This perspective was presented by Dawkins and Krebs (1978; Krebs & Dawkins 1984) in a pair of seminal papers which suggested that in situations in which the interests of individuals differed, communication could not be understood as the honest exchange of information for the greater good, but must be characterised as competitive exploitation. This exploitation could occur in two forms. Signallers might exploit receivers, manipulating them with behaviour which receivers would incorrectly interpret[2]. Conversely, observers might exploit signallers, "reading their

---

[2] N.b., This use of the term 'interpret' should not be taken to imply any particular cognitive ability on the part of signal observers. Rather, an observer's interpretation of a signal should be read as shorthand for an observer's response to a signal.

minds" through paying attention to some inadvertent slip which betrays their intent (or some other information of interest to observers).

The authors proposed that these twin factors might entwine in an escalating series of adaptations on the part of signaller and observer, leading to ever more devious forms of exploitation and counter-exploitation — an evolutionary arms race (Dawkins & Krebs, 1979).

From this perspective, organisms with no conflict of interest could still be characterised in the terms used by previous theorists. Such organisms are subject to the selective pressures favouring efficiency described in the previous section. As such they are hypothesized to enjoy a co-evolutionary trajectory which leads to ever more efficient signallers and ever more attentive observers, as evolution adapts both to effectively exchange information in order to further some joint project. Thus, co-operative scenarios are predicted to lead to stable, quietly effective systems ("conspiratorial whispers"), whereas competitive scenarios are predicted to lead to insecure, stridently aggressive systems which threaten to collapse at any instant.

Many natural signalling scenarios have been identified as involving the potential for conflicts of interest, including parent-offspring communication (Godfray, 1995), predator-prey signalling (Endler, 1991), aggressive communication (Maynard Smith, 1974b), and courtship displays (Trivers, 1972). Indeed, few signalling scenarios (outside those involving eusocial insect societies) can be classed as truly co-operative, i.e., involving no potential for a conflict of interests between signaller and receiver. As a result, under the reading provided by Krebs and Dawkins, most natural communication must be regarded as unstable. This conclusion has led to deep problems with the conceptual vocabulary which had previously been employed unproblematically in the description of signalling behaviour.

For example, in situations in which observers exploit tell-tale signs which inform them of signaller characteristics, in what sense is the signaller actually signalling? Surely if a vulture spots some carrion which is as yet undetected by its conspecifics (with whom it competes for food), and descends from the sky to devour said carrion, the act of descending, as seen by other vultures, cannot be taken as a signal from the first vulture to its competitors to the effect that 'there is some food to be eaten over here'? In what sense would one want to class the first vulture as a signaller in this scenario?[3]

Similarly, if a signaller exploits an observer's sensory bias, in what sense can the observer be said to have 'observed a signal'. For example, consider a species of aquatic bacteria indigenous to the waters of the northern hemisphere, for which oxygen is toxic. Magnetosomes within these bacteria pull them towards magnetic north in order to direct them away from the surface of the water, thus away from oxygen. If a certain aquatic predator of bacteria could generate an electro-magnetic field about itself which exploited the directional system within its prey such that it steered them towards it, in what sense is this magnetic field acting as a signal?

The root of this conceptual morass has been identified as the role of information within signalling systems (e.g., Markl, 1985; Hasson, 1994; Dawkins, 1995; Maynard Smith & Harper, 1995). The term information may be used in its technical sense (Shannon & Weaver, 1949) in which it means a reduction in uncertainty on the part of a signal receiver, or a more colloquial sense in which it means something along the lines of an increased ability to effectively decide

---

[3] Thanks to Jason Noble for this example.

between courses of action, or simply an increase in relevant knowledge. Information exchange, in both of these senses, might not occur within animal communication, given the characterisation of signalling provided by Dawkins and Krebs (1978). This leads them to suggest that, for the purposes of describing animal communication, one should "abandon the concept of information altogether" (p.309).

In attempting to assess the role of information in the definition of communication, theorists run into a problem. If signals are defined as essentially informative then lies, cheats, bluffs, etc., cannot be signals since, by definition, they are not informative. However, if signals are defined with no mention of their role as information carriers, might not camouflage, crypsis, or even hiding fall under the category of signalling behaviours? To see how this might be the case, consider a definition of signalling which is prevalent in the literature (e.g., Wiley, 1983; Endler, 1993; Johnstone, 1997).

If signalling occurs when 'the behaviour of one organism causes a change in the behaviour of another organism which results in a fitness increase for the first organism', then the bacterial predator alluded to above must clearly be classed as a signaller (although the vulture cannot since it loses out to its competitors through giving away the location of a valuable resource). Furthermore, the avoidance of predation through camouflage, crypsis, or hiding, all result in increases in fitness due to the 'manipulation' of predator behaviour, and thus all must be classed as communication under such a definition.

In addition to the terminological problems engendered by the exploitative conception of signalling proposed by Dawkins and Krebs, the absence of a role for veridical information within their account results in an attendant lack of evolutionary stability within their portrayal of communicative systems. Maynard Smith and Harper (1995, p.305) state this explicitly when they contend that "there is a connection [between information and evolutionary stability] because it is not evolutionarily stable for the receiver to alter its behaviour unless, on average the signal carries information of value to it".

In the remainder of this chapter, discussion will focus upon three conditions in which evolutionary stability, or at least partial stability, are hypothesised to arise despite the conflicting forces identified by Krebs and Dawkins. The first of these conditions results in what will be termed *conventional signalling* since the stability of this class of signalling is maintained by the observation of a convention. Subsequently, *sensory exploitation*, and *handicap signalling* will be presented as developments of the notion of conventional signalling. The validity of these two hypothetical classes of signalling will be assessed, and attention will be drawn to aspects of the theory within each class which will be addressed by models within chapters 6 and 7, respectively. Subsequently, a characterisation of signalling which draws heavily on Millikan's (1984, 1993a) work within a philosophical field she dubs *teleosemantics* will be presented in an effort to highlight the difference between the three classes of signalling presented herein.

5.3 Conventional Signalling

An example of a conventional signal is the use of a red light to indicate to motorists that they are required to stop, and a green light to indicate that they are permitted to go; this is the convention observed on British roads. The signal is conventional because there is an arbitrary but consistent relationship between signal form and signal meaning. Although there may be some historical reason for the assignment of red and green, such reasons are mere accidental contingencies from the perspective of the contemporary motorist. It is conceivable that the British road-traffic system could have come to use green to mean stop and red to mean go without rendering the convention unworkable.

Conventions such as those observed by motorists are enforced by society through the application of law. If there are conventional signalling systems within nature, how might they be policed?

First, co-operative signalling systems may be excluded from the discussion. The bee dance, a classic example of a natural signalling system, needs no ‘policing’, since there is no threat of insurgent deception. The reproductive success of each foraging bee is coincident. Therefore any advantage that one bee might contrive to achieve over another is countered by an equal disadvantage through the reduced kin-selected element of its inclusive fitness (indeed, foraging bees have no other component of inclusive fitness since they leave no offspring themselves).

Signalling under such co-operative situations will be termed *co-operative signalling*. The characteristics of co-operative signals may well be contingent, arbitrary, etc., but the term conventional signalling will be reserved for signalling systems which feature arbitrary signals, *and* demand some regulation to prevent the occurrence of cheating.

However, although co-operative systems are inherently honest, they may provide the preconditions for conventional signalling. For example, aposematic prey which advertise their unpalatability with gaudy colouration, or unpleasant smells, are involved in a co-operative signalling venture with predators which attend to such signals and preferentially predate prey items exhibiting no such warnings. Both parties stand to gain from this practice, and, as such, warning signals and avoidance behaviour would remain stable but for the occurrence of mimics.

Mimics are best conceived of as third parties, parasitising a co-operative signalling system. They exploit the selected-for tendency amongst predators to avoid prey with warning colouration by adopting the signal without investing in unpalatability. However, although it is possible that mimics might destabilise a co-operative signalling system, a regulatory pressure exists in the form of negative feedback on the frequency of mimicry within the signalling system. As the frequency of mimics increases within a signalling system, the value of the warning signal gradually degrades. The correlation between signal and unpalatability weakens. As this occurs, the worth of avoidance behaviour also degrades, promoting an increase in the fitness of predatory strategies which ignore warning signals. This in turn decreases the fitness of mimics.

Thus as the frequency of mimicry increases the fitness of this strategy decreases. Perhaps the frequency of mimicry within a conventional signalling system will oscillate continually, or reach a stable level at which a low proportion of mimics co-exist alongside a co-operative signalling system, or perhaps mimics will repeatedly appear, increase in numbers briefly, and subsequently go extinct.

A fourth option exists in the possibility that as mimics increase in frequency, the co-operative signalling system that they exploit will collapse, or adopt a new, as yet unparasitised, convention with which to signal aposematism. Andersson (1980) suggests that this possibility may account for the high number of different threat displays utilised by individual species. Evolutionary simulation models of conventional signalling seem well suited to the task of discovering the facts which govern the character of the co-evolutionary dynamics which are involved within such signalling systems.

An alternative, and complementary, mechanism which results in conventional signalling arising from co-operative signalling scenarios involves the intrusion of a third party which exploits the honest *signaller*, rather than receiver. This parasitic strategy might take the form of "eavesdropping" predators which orientate towards prey signals.

From this perspective conventional signalling may only arise as the result of exploitation of a prior co-operative signalling system. Indeed, a definition of a conventional signalling scenario may be stated as a co-operative signalling scenario suffering exploitation which is regulated through a negative feedback process, where co-operative signalling is defined as signalling between individuals which experience no conflict of interest (i.e., individuals which rank the possible outcomes of a signalling interaction identically).

5.4 Sensory Exploitation

Although Krebs and Dawkins' (1984) characterisation of manipulation and mind-reading accentuates the unstable character of the signalling systems which result from these two exploitative processes, more recent authors have stressed that limiting factors may prevent the predicted arms races from escalating unchecked.

Although some traits exploited in signallers or observers may be free to evolve under the selective pressure that this exploitation results in, the evolution of other traits may be constrained by factors external to the signalling context. Ryan (1990, 1997; Ryan & Rand 1993) suggests that sensory biases within an observer which are being exploited by signallers may resist evolutionary change for two reasons. First, such sensory biases may be pre-existing adaptations selected for in contexts other than that of the signalling scenario. Their adaptive value within these non-signalling contexts may preclude their exaptation. Secondly, although such sensory biases may have no adaptive value *per se*, they may resist evolutionary change as a result of their being the necessary side-effect of some associated adaptive trait. In such a situation, the exploited trait may be very hard to select out of the observer population. For example, if, as Gould and Lewontin (1979) contend, the human chin serves no purpose, but is merely contingent upon several facial adaptations which indirectly influence its character, selective pressure to alter the chin's character may prove ineffective since the changes in facial structure which must necessarily accompany such a change in chin characteristics might militate against such evolutionary change.

An associated formulation of the sensory bias, or sensory trap, hypothesis is provided by Arak and Enquist (1993) in terms of 'hidden preferences' which are postulated to be inherent in many sensory systems. The authors point out that the responses of sensory systems to supernormal stimuli (i.e., stimuli which are not part of the selective history of the sensory system) are free to drift and may exhibit a phenomenon known as 'peak shift' in which stimuli outside the range of normal experience elicit exaggerated responses. Such hidden preferences are ripe for exploitation by signallers.

However, during such a period of exploitation, as stimuli which would previously have been classed as supernormal become frequent, the exaggerated response that they elicit may be selected against. Without an independent selective pressure in favour of the peak shift effect, over evolutionary time its strength will tend to reduce until it disappears, and with it the ‘hidden preference’ and the attendant exploitation.

Recent attempts to model the evolution of signalling systems as the exploitation of sensory biases have included accounts of the evolution of complex symmetry, and the evolution of exaggerated traits, which rely on neural networks, and artificial evolutionary algorithms (e.g., Arak & Enquist, 1993, 1995a; Enquist & Arak, 1993, 1994; Johnstone, 1994a; Hurd, Wachtmeister, & Enquist, 1995).

One of these models (Enquist & Arak, 1994) will be explored in the next chapter. An attempt will be made to demonstrate that the argument advanced in order to explain the complex symmetrical form of many animal displays is flawed, and to explore the methodological problems with the model which led to this flawed conclusion.

A complementary class of stable signalling scenarios may be constructed as a corollary of the reasoning presented above in the form of signalling systems which exploit persistent cues. In such a system, observers exploit a behavioural regularity, or motor bias, in signaller behaviour which allows observers to accurately derive some useful information. Signallers persist in providing this cue, despite the selective pressure to resist exploitation through exaptation of this behavioural correlate, due to the adaptive significance of the behaviour in alternative contexts, or the difficulty of selecting out the particular motor bias whilst maintaining general viability.

As a pair, the exploitation of sensory biases, and the exploitation of motor biases, will be referred to as *exploitative signalling*. It should be noted that whereas conventional signals are typically described as being of arbitrary form, exploitative signals are constrained by the context within which the bias which they exploit was originally selected for.

For example, if the visual system of a species has evolved to preferentially attend to areas of the visual field which are a certain colour (e.g., red) since prey items are typically red, this bias might be exploited by suitors during courtship displays in order to gain the attention of the selecting sex. If the colour of prey types changes over evolutionary time, the predator’s associated sensory bias may well adapt by tracking this change in prey colouration. As a result, courtship displays might track prey colouration in order to remain eye-catching.

Although a similar conventional, or co-operative, signal along the lines of ‘Look-at-me-I-am-soliciting-a-copulative-episode!’ might similarly exploit a latent preference within the observer’s sensory system initially, such signals are free to evolve subsequently in an essentially arbitrary manner (although this evolution may be constrained by the nature of the medium through which the signal is passed, and the morphological and behavioural constraints of both signaller and receiver).

Thus, within a signalling system in which a sensory bias is being exploited (one that has stabilised despite this exploitation for the reasons discussed above) the form of the signal is less arbitrary than those found within conventional, or co-operative, signalling systems, since the historical contingencies which governed its inception actively constrain its subsequent form.

5.5 Handicaps

An alternative notion, dubbed the ‘handicap principle’, was proposed by Zahavi (1975, 1977a, 1977b) with the intention to provide a mechanism by which stability might be enforced within natural signalling systems. The thrust of Zahavi’s argument is reminiscent of the notion of *con-*

spicuous consumption discussed by Veblen (1899).

As part of his theory of the ‘leisure class’, Veblen contends that over-indulgence in eating, drinking, and consumption in general has become a sign of societal status. One’s wealth or rank can be ascertained by attending to the extent of one’s “unproductive consumption”, that is one’s consumption above and beyond that necessary to sustain oneself. Indeed, Veblen goes so far as to suggest that the root of this association is embedded in mankind’s evolutionary prehistory, in “the beginnings of predatory life”.

The logic behind the use of conspicuous consumption as an index of social status turns on the observation that those of lower social standing cannot afford to over-consume; their scant resources are more than accounted for by the continual demands which must be met merely in order to survive. High ranking socialites, on the other hand, may demonstrate their membership of the leisure class through freely wasting resources on goods which are expensive to procure, or require prohibitive amounts of free time to enjoy. Furthermore, the extent of one’s wealth will tend to be correlated with the degree of extravagance which one evidences, allowing conspicuous consumption to distinguish, not merely between members of the working and leisure classes, but also between members of the leisure class who differ in wealth.

Zahavi has suggested that a similar mechanism is at work within natural signalling systems, ensuring that advertisements are costly and honest. However, whereas Veblen clearly classes the phenomenon of conspicuous consumption primarily as a convention maintained by elaborate social customs, taboos, tacit understandings, and rules of conduct, Zahavi views the handicap principle as a law, governing behaviour rather than merely describing it.

In his initial papers, Zahavi (1975, 1977a, 1977b) suggested that honesty could be maintained within a signalling scenario involving a potential conflict of interest between signaller and receiver as a result of the cost that signallers incurred in producing a signal. Although Zahavi gave examples from many different classes of scenario (e.g., mate choice, parent-offspring conflict, warning signals, threat displays, altruistic signalling, Zahavi 1977b), here the argument will be cast in terms of sexual selection.

Zahavi maintained that the degree of cost experienced by a suitor when making a particular signal might be negatively correlated with the viability of the suitor. Since the viability of a potential mate is of interest to a choosy female, she should be selected to take advantage of any honest indicator of this male trait. In such a situation, Zahavi claimed, signallers would come to advertise at a level which correlated with the information of interest to observers. Such advertisements would be honest, and could not be exploited by bluffers, since a bluffer (a signaller which produces a larger advertisement than an honest signaller would produce under the same conditions), could not afford to pay the increased cost that such a bluff would entail.

The cost of such signals was characterised by Zahavi as a ‘handicap’, since it is through demonstrating its ability to bear such a handicap that a signaller can convince an observer that it is telling the truth. He proposed that his ‘handicap principle’ could account for the extravagant signals used in many signalling contexts. For example, for Zahavi, peacock trains (tail feathers) — the quality of which is a factor governing female mate choice (Petrie, Halliday, & Saunders, 1991) — are handicaps. Peacocks bearing large trains suffer energetic costs in producing and maintaining such a train, costs in terms of an increased risk of predation, etc. Only those peacocks with a higher than average viability will be able bear the costs associated with a more elaborate than average train (see Petrie, 1994, for empirical data which supports this claim). Thus peahens can accurately gather veridical information concerning mate viability by attending to the quality of a male's display.

As this verbal argument stands, it is hard to see clearly whether handicaps can remain evolutionarily stable. For example, if females select males with high viability on the basis of the extravagance of their courtship display, might the fact that their offspring inherit both high viability *and* a costly handicap preclude the stability of such signalling systems (Maynard Smith, 1976)?

Zahavi has subsequently developed the handicap principle, constructing an elaborate and general theory of what he terms *signal selection* which is intended to account for most, if not all, signalling behaviour — from inter-cellular signals and allofeeding in avian social hierarchies, to human suicide attempts and method acting (Zahavi, 1980, 1982, 1987).

Zahavi parallels his theory of signal selection with those of natural selection and sexual selection, going so far as to suggest that signal selection should replace sexual selection as an explanatory mechanism capable of accounting for the phenomena associated with female choice and exaggerated signalling (Zahavi, 1991).

Zahavi's intention is thus clearly to provide a near universal mechanism with which to account for most features of signalling behaviour, both within sexual contexts, and alternative contexts involving a conflict of interest between signaller and receiver. Within chapter 7 herein the potential for such ubiquity will be assessed.

However, here we shall concern ourselves solely with Zahavi's original account of the handicap principle, and subsequent attempts to formalise and explore the validity of this mechanism. Zahavi's (1975, 1977a, 1977b) verbal argument was originally presented as a series of descriptions of natural signalling scenarios, each cast as an example of the handicap principle at work. For each signalling system, signalling costs were diagnosed, and links between these costs and the quality being advertised were identified. Once this had been accomplished, the signals could be cast as handicaps, honestly revealing some information of interest to observers. In constructing this 'argument by example', the different types of signalling cost invoked, and the variety of ways in which cost was linked to quality, provided theorists with a spectrum of interpretations of the handicap mechanism, each of which demanded a formal treatment.

5.5.1 Genetic Models of the Handicap Principle

As a result of the multiple possible interpretations of Zahavi's verbal arguments, theorists proceeded to formalise the handicap principle in a variety of ways. A number of these early models failed to support Zahavi's counter-intuitive notion that wasteful extravagance might make evolutionary sense (e.g., Davis & O'Donald, 1976; Maynard Smith, 1976, 1978, 1985; Kirkpatrick, 1986). However, alternative models were constructed which provided some support for the notion that at equilibrium signals might be both costly and honest (e.g., Eshel, 1978; Andersson, 1982; Hamilton & Zuk, 1982; Nur & Hasson, 1984; Pomiankowski, 1987).

With hindsight, this period appears to be one of gradual revision and consolidation of competing interpretations of the handicap principle. Maynard Smith (1985) identifies three such contrasting interpretations.

The first account of the handicap principle, termed the pure epistasis account by Iwasa, Pomiankowski, and Nee (1991), allows that a costly ornament reduces the fitness of low quality males to a much greater degree than that of high quality males. If, at equilibrium, low-quality males do not possess the ornament, but high-quality males do, then females choosing ornamented males will mate more frequently with high-quality males than those of low quality.

Conditional handicap accounts (e.g., West-Eberhard, 1979) suggest that if whether an expression is expressed or not is dependent on signaller condition, which is itself correlated with a trait of interest to observers, such observers might benefit from utilising the ornament as an advertisement of the relevant signaller trait.

Revealing handicaps (e.g., Hamilton & Zuk, 1982) are ornaments, the condition of which displays some quality of interest to observers. For example the condition of male display plumage might enable females to assess the parasite load of a potential mate. However, females solicited by a suitor without such special plumage would be unable to judge his parasite load. In such circumstances females might be selected to accept only males with display plumage, and to preferentially select those within this group which were judged to have low parasite loads.

Genetic models of all three classes of account have produced varying results, but generally appear to offer little support for the handicap principle. Such models, after critical examination, demonstrated that, even if handicap signalling were stable, it could not be reached, over evolutionary time, by a population in which handicaps, and a preference for handicaps were both initially rare (see Maynard Smith, 1985, for a review of these early models).

5.5.2 Game Theoretic Accounts of the Handicap Principle

The first attempt to model the handicap principle in terms of ESS theory was provided by Enquist (1985). In an important paper, Enquist demonstrated that signals of aggressive intent could be honest only if they were risky to make. This result is the forerunner of a prolific and successful literature which casts the handicap principle in terms of strategic signalling and its associated costs. However, this model, and most of those which followed it involve two limiting assumptions.

First, Enquist's model assumed that the quality being advertised by a signaller was environmentally determined rather than inherited. Enquist specifies that a population engaged in the signalling game he defines always consists of equal proportions of weak and strong players. Successful players will leave offspring which inherit their parents' strategy, but not their parents' quality. Second, Enquist's model does not address the problem of whether evolutionary trajectories exist which lead from populations which do not use handicaps as honest signals to populations which do use handicaps in this manner. Although a few of the subsequent models do address these issues (e.g., Yachi, 1995, explores the conditions under which honest handicap signalling may evolve from an initially non-signalling population) they remain relatively under-investigated.

ESS models of aggressive contests had been constructed prior to Enquist's model (e.g. Maynard Smith & Price, 1973; Maynard Smith, 1974b; Maynard Smith & Parker, 1976), and had been the subject of debate concerning the degree to which such models could contribute to, and accommodate, ethological studies of the same phenomena (Caryl, 1979, 1981, 1982; Hinde, 1981). However, Enquist's result is rightly recognised as motivating a new perspective on the handicap principle which led to what is regarded as the first conclusive formal demonstration of the validity of Zahavi's handicap principle (Grafen, 1990a, 1990b).

Grafen cast the handicap principle in terms of sexual selection. He then proceeded to show that in order that an advertisement of male quality be honest, it must be costly, and that the cost of the advertisement must decrease as male quality increased. That these two conditions must be met in order that honest signalling can be evolutionarily stable is now recognised as the core claim made by Zahavi in his original papers, although the extent to which Zahavi himself recognised the centrality of these two conditions is debatable (Iwasa et al., 1991).

Subsequent to Grafen's complex, continuous model, several authors have provided corroborative models which seek to simplify or extend his work.

Maynard Smith (1991) provides a simplified, discrete model of the handicap principle, termed the Philip Sidney game, in which he demonstrates that the signalling of need between relatives may be honest, and thus stable, only if the value of the resource differs with need, and the production of the begging signal is costly. A continuous version of this model has been provided which supports its original findings (Johnstone & Grafen, 1992a)

Godfray (1991) provides a similar result to that of Maynard Smith (1991) in demonstrating that the honest signalling of offspring need to parents, in order to solicit parental resources, is stable if (i) signals are costly, and (ii) the value of parental resources increases with offspring need.

Pomiankowski, Iwasa and Nee (1991; Iwasa et al., 1991) model a signalling scenario in which, unlike previous game theoretic models, the advertised trait is heritable. The authors demonstrate that the paradox of the lek (Kirkpatrick & Ryan, 1991; Pomiankowski & Møller, 1995) can be avoided if deleterious mutation events are biased such that they maintain variability in the advertised trait. Under such conditions honest signalling of underlying genetically specified viability may be stable despite observers paying costs for choice.

The model is pitched in terms of an additive quantitative genetic model in which correlations between traits are represented as additive genetic co-variances. Although the model is admirably concise, the assumption that certain co-variances remain positive over evolutionary time appears suspect (Andersson, 1982). Evolutionary simulation models in which these postulates might be allowed to vary (i.e., to become observable quantities rather than specified quantities) seem ideal tools with which to address such concerns.

Johnstone and Grafen (1992b, 1993; Grafen & Johnstone 1993 ; Johnstone 1994) have explored the addition of noise to Grafen's (1990b) error-free, continuous model of honest signalling. They demonstrate that uncertainty on the part of receivers can lead to stratification of the signalling population into a series of discrete groups of signallers, each defined by a range of signaller quality, emitting an identical signal, and emitting signals which increase with the mean signaller quality within the group. This stratification appears to result from the fact that observers coping with error-prone signallers tend to ignore small differences in signal strength, since such differences may arise as the result of noise within the system. Signallers who might honestly signal just below the thresholds imposed by observers gain no benefit from signalling any stronger than those slightly poorer than themselves, and will tend to adopt a slightly attenuated signal as a result. This process is sufficient to account for the bands of signallers within which, despite varying in quality, signallers are selected to signal identically.

Vega-Redondo and Hasson (1993) apply the handicap principle to predator-prey signalling in which prey indicate their ability to outrun a predator. The authors relate their results to parameters of their model, including herd size, the range of distinguishable signals available to prey, and the distribution of prey quality. They conclude that the handicap principle can account for the different signals given by prey to predators, with higher-quality prey giving riskier signals than their low-quality conspecifics.

Maynard Smith (1994) extends a result from his previous Philip Sidney game, which demonstrated that cost-free signals could be stable if signallers and receivers ranked possible outcomes in the same order (i.e., if signallers and receivers did not experience a conflict of interest). Here the author extends this reasoning to a continuous game, demonstrating that the same result will not hold if quality is distributed uniformly within both signalling, and receiving populations. Candidate distributions which do allow stable, cost-free signalling are presented, and the likelihood of such distributions characterising natural signalling systems is discussed.

Adams and Mesterton-Gibbons (1995) draw a distinction between production handicaps, for which handicap costs are paid in order to produce the signal (e.g., acoustic displays Burk, 1988), and vulnerability handicaps, for which costs are only paid if cheats are probed (i.e., their bluff is called). The authors point out that Grafen's (1990a, 1990b) model, and models which follow it, assume that the value of an observer response is equal for all signallers, irrespective of signaller quality, or that this value increases with signaller quality. They contend that, although this may be the case in some signalling scenarios, aggressive signalling between belligerents prior to combat may involve benefits which are greater for weaker signallers than for their stronger competitors. For example, the benefit of avoiding a violent battle is probably greater for the weaker of a pair of antagonists since the weaker of the pair is most likely to suffer a potentially life-threatening wound should such a battle take place.

The authors proceed to model such contests, demonstrating that under certain conditions evolutionarily stable signalling may involve threatening signals on the part of the weakest, and strongest, individuals, and no signalling on the part of individuals of intermediate strength. The degree of bluffing (weak individuals who threaten to escalate a contest) is related to the higher benefits such individuals obtain when they are believed, and the fact that they only pay a handicap when probed. The authors note that if the handicap were modelled as a production cost, no such ESS would exist. This is an interesting but complex model. Similar notions to those it addresses will be explored in chapter 7 herein.

Hurd (1995, 1997) constructs a very simple, discrete, action-response game featuring a signaller and receiver who experience a conflict of interests. The author demonstrates that Zahavi's conditions for honesty are sufficient, but not necessary, for honest signalling equilibria to exist, and further points out that honest signalling may be stable in situations within which high quality signallers pay no costs for making an advertisement of their quality (or even preferentially make this advertisement rather than some other behaviour) as long as the cost of such an advertisement for low-quality signallers is prohibitively large. This model will be extended in chapter 7.

Krakauer and Johnstone (1995) provide a model which attempts to reconcile the handicap principle with conventional signalling. Employing artificial neural networks as models of simple organisms involved in a signalling scenario, the authors demonstrate that handicap signalling is relatively stable, but is susceptible to occasional bursts of dishonesty. Such dishonesty is suppressed through negative feedback interactions typical of conventional signalling scenarios.

Krakauer and Pagel (1995) provide an evolutionary simulation model based upon game theoretic considerations, in which they demonstrate that a spatially distributed population of signallers and receivers may pay no signalling costs, yet maintain an appreciable degree of honest communication for long periods of evolutionary time. Honesty is stable as a result of the variation in strategy densities which necessarily occurs across spatially distributed populations, which ensures that in certain areas the frequency with which honest individuals encounter one another may be higher than that with which they encounter deceivers, and that therefore islands of honesty may exist for long periods within a sea of dishonesty.

Dawkins and Guilford (1991) have drawn attention to the role played by receivers in signalling systems. The authors highlight the effects of receiver costs on the stability of handicap signalling. Receivers pay costs in terms of wasted time, increased risks of predation, or infection. Within signalling systems in which receivers and signallers alternate roles (e.g., aggressive signalling by stags Clutton-Brock & Albon, 1979) each must pay the costs of signalling in the current bout, in order to gain information during the next bout. The authors suggest that, for scenarios in which costs are involved in the receipt of signals in addition to their production, receivers may settle for less than perfect honesty as a means of reducing the costs associated with their role within the signalling system to a strategically acceptable level. For example, if attending a lek in order to distinguish between suitors of differing quality involves a risk of predation, or infection, females may spend less time choosing between males than would be necessary in order to ascertain their quality beyond doubt. Females might spend just enough time to make a judgement which ensured that the *benefits* of mate selection balanced the *costs* involved in mate selection.

This brief review of handicap models is by no means exhaustive, but serves to demonstrate the variety of phenomena explored under the rubric of handicap signalling, and the potential for elaboration, and extension which is inherent within the handicap mechanism.

The models and arguments sketched above may be divided into categories. The first category contains those models which attempt to demonstrate that signaller *quality* may only be advertised honestly under the conditions specified by Zahavi. This category includes Enquist's (1985) discrete model, and Grafen's (1990a, 1990b) continuous model. The second category contains those models which attempt to demonstrate that signaller *need* may only be advertised honestly under the conditions specified by Zahavi. This category includes Maynard Smith's (1991) discrete model, and Godfray's (1991) continuous model.

Johnstone (1997) has described these two categories, both of which stipulate that signalling must carry a cost for signallers, whereas observer responses must benefit signallers. The first category of models have led theorists to conclude that the honest advertisement of signaller quality may only be evolutionarily stable if signalling costs differ with quality such that higher quality signallers pay less for any given advertisement than their poorer conspecifics. The second category of models have led theorists to conclude that the honest advertisement of signaller *need* may only be evolutionarily stable if the benefits of observer response for the signaller differ with signaller need such that more needy signallers gain more from a given observer response than their less needy conspecifics.

A third category of model involves the addition of some feature to a model in which handicap signalling has been shown to be stable. This additional feature typically reduces the degree of honesty within a system, but does not totally destabilise it. For example, the addition of noise by Johnstone and Grafen (1992b), or the addition of receiver costs by Dawkins and Guilford (1991), serve to inject a degree of deceit or misinformation into an otherwise honest signalling system.

In chapter 7, the distinction between the first two categories will be challenged. Why should signalling costs vary with quality but not with need? Why should the benefits of observer response vary with need, but not with quality?

5.6 Defining Communication

As was mentioned at the outset of this chapter, there have been frequent attempts within the animal signalling literature to define communication in a manner which allows morphological and behavioural traits that function as signals to be distinguished from those that do not (e.g., Hasson, 1994; Maynard Smith & Harper, 1995).

A brief discussion of a definition of signalling will be presented here with an aim of providing a consistent terminology with which to support the definitions of *co-operative*, *conventional*, *exploitative*, and *handicap* signalling offered above. The construction of this terminology will closely follow Millikan's (1984, 1993a) development of an evolutionary-functional account of mental content. The utility of Millikan's conceptual scheme has been promoted within behavioural ecology (Bekoff & Allen, 1992), but has so far received little attention. Here, a very minimal account of her distinction between signals (intentional icons in her terminology) and tacit suppositions will be used to distinguish the four classes of signalling defined above.

Consider a scenario, depicted schematically in figure 5.1, in which there exists a certain state of affairs, Q, to which a signaller, S, is privy, but which is unknown to an observer, O. Furthermore, accurate knowledge of Q would benefit O. S is built such that it makes an action, A, whenever Q obtains. O is similarly built such that it makes a response, R, whenever A obtains. Furthermore, assume that R and A have fitness implications for S and O, and that these fitness implications might be dependent on Q.

For example, S and O are engaged in a game of poker. A degree of escalated betting has occurred, and a considerable pot is at stake. Take Q to indicate that S possesses a high quality hand. Action A corresponds to a facial twitch on the part of S, whilst response R corresponds to O folding her hand.

Is this a communicative scenario? As yet we cannot be sure.

For many biologists the matter would turn on the character of the fitness implications of A and/or R for S and/or O. As has been mentioned above, one definition of communication (Wiley, 1983; Endler, 1993; Johnstone, 1997) demands that A must alter O's behaviour such that S enjoys an increase in fitness. However, it has been pointed out that such an increase in fitness might occur fortuitously, or as the result of some consequence of the constitution of S, and/or O which was not selected for (or which was selected-for outside the signalling context). As discussed above, hiding behaviour on the part of S might increase S's fitness as a result of its 'effect' on O's behaviour. However, hiding is not naturally understood as 'signalling that one is not there'.

Some theorists (e.g., Maynard Smith & Harper, 1995) have recognised this problem, and have

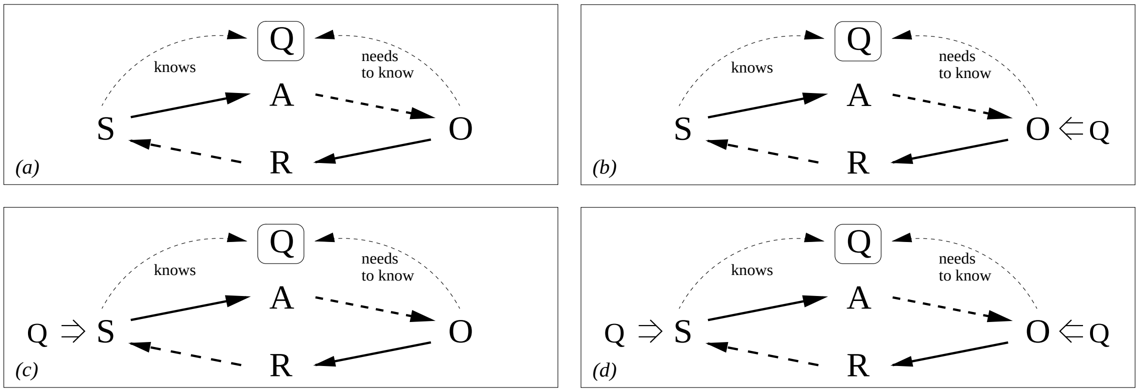


Figure 5.1: Schematic depictions of dyadic interactions which potentially constitute signalling scenarios. Each depicts a state of affairs, *Q* (in a window near the top of each frame), in which a signaller, *S*, always makes an action, *A*, to which an observer, *O*, always makes a response, *R*. Behaviours *A* and *R* have fitness consequences for *S* and *O* which may depend on *Q*. Solid arrows intervene between actors and actions, dashed arrows intervene between actions and recipients, arcs indicate the relationship between an actor and the state of affairs, *Q*, whilst ⇒ and ⇐ intervene between a state of affairs and an actor which has been selected to act in a certain way under that state of affairs. *(a)* The behaviours of *S* and *O* are unselected-for, or have been selected-for outside the signalling context. *(b) O* has been selected to exploit *S*'s motor bias, i.e., mind-reading *sensu* Krebs and Dawkins (1984). *(c) S* has been selected to exploit *O*'s sensory bias, i.e., manipulation *sensu* Krebs and Dawkins (1984). *(d) S* has been selected to produce behaviour *A* under state of affairs *Q*, *and O* has been selected to make response, *R*, when in receipt of action *A*.

demanded that *O*'s behaviour change as a result of *S*'s action such that both *S*, *and O*, enjoy an increase in fitness. For example, the dance performed by a returning bee which informs its conspecifics of the location of nectar (when performed correctly) results in fitness benefits for all bees concerned.

However, consider a situation in which a bee has discovered a source of nectar and is returning to the hive in order to advertise the location of this source of nectar. However, unbeknownst to the returning bee, this nectar has been consumed, or moved to a different location, during its hive-bound fight. Under such conditions, the dance performed by the returning bee results in a fitness decrease for all bees concerned, yet it would seem churlish to deny that the bee was signalling in such a case. Indeed the potential for bee dances to fail to veridically describe the location of nectar is identified by Maynard Smith and Harper (1995) as a primary characteristic of signals, when they quote Eco's (1976) description of semiotics (the study of signs) as "the discipline studying everything that can be used in order to lie".

Accordingly, Maynard Smith and Harper (1995) refine their definition to read something like 'a signal is a morphological or behavioural trait which elicits a response in an observer that, on average, increases the fitness of signaller and receiver'.

What is the meaning of 'on average' here? Fifty percent of the time? Most of the time? Enough of the time? Millikan (1993b) freely admits her ignorance of the exact frequency with which bee dances accurately map the location of nectar. She is unconcerned with this exact frequency since the rate at which such dances correspond accurately to the location of nectar merely has to be sufficient to ensure a selective advantage in those hives which employ such dances. That a bee dance behaviour is accurate enough of the time is evidenced by the evolutionary success of such a behaviour (i.e., the continued presence of dancing behaviour within a bee species).

This understanding of 'on average' seems to agree with the spirit in which Maynard Smith and Harper (1995) employ the term. In order that signalling be stable, there must be a selective advantage in attending and responding to signals; similarly, there must be a selective advantage in making them.

There are at least four different situations which might be described by the schematic scenario involving *S* and *O* as it has so far been cast. First, *S* and *O* might produce behaviours *A* and *R*, respectively, under circumstances *Q*, as a result of some unselected-for side-effect of their constitutions, or an aspect of their constitutions selected for in a context other than poker playing (see figure 5.1a). In such a case, *S*'s tendency to twitch when holding a high quality hand, and *O*'s tendency to fold her hand when in the presence of such a tic, are either non-adaptive spandrel-like behaviours (Gould & Lewontin, 1979), or are, at the very least, non-adaptive traits with respect to the poker game.

Scenarios involving such traits cannot count as communication, despite the facts that (i) in these scenarios the presence of action A correlates perfectly with the presence of state-of-affairs *Q*, and (ii) in some scenarios of this type the actions of one individual (*S*) affect the behaviour of another (*O*) in a manner with positive fitness consequences for the signaller, receiver, or even both. Although it may be the case that both *S and O* benefit in fitness terms from their behaviour, since these behaviours were selected for in alternative contexts, it seems unnatural to term them 'signalling' and 'observing', despite the fact that such behaviours are stable, and beneficial.

The second and third relevant situations are complementary. In each, the behaviour of one member of the dyad has been selected for in the context of poker playing, whilst the behaviour of the remaining member has not. For example, figure 5.1b depicts a scenario in which, whilst *S*'s twitch is an unselected-for muscular spasm (or a muscular spasm adapted to some alternative context), *O*'s response has been selected for on the basis that ancestral *O*'s which threw in their hand when ancestral *S*'s evidenced a facial twitch did better at poker than their competitors (ancestral *O*'s playing alternative strategies).

Similarly, figure 5.1c depicts a scenario in which *O*'s response to *S*'s facial twitch is an unselected-for side-effect of *O*'s general psychology (or was selected for in a context alternative to poker playing), whilst *S*'s twitch has been selected for on the basis that ancestral *S*'s employing such a twitch outperformed their competitors.

The former of these pair of complementary situations corresponds to 'mind-reading' in the terminology of Krebs and Dawkins (1984), whilst the latter corresponds to 'manipulation' (Dawkins & Krebs, 1978; Krebs & Dawkins, 1984). Similarly, the exploitation of sensory biases (Ryan, 1990), and the reading of cues (the exploitation of motor biases) fall under these two categories.

For Millikan, neither of these scenarios involves signalling. The behaviour of one party is merely adapted to exploit the character of the other. Millikan (1993b) characterises such adaptations as involving "tacit suppositions". For example, although the circadian rhythms of many animals are clearly linked to the speed of the earth's rotation, such rhythms do not *represent* the earth's angular velocity. Similarly, although a chameleon's skin matches the colour of its surroundings the skin does not *signal* anything.

Both the systems which govern circadian rhythms and the systems which govern chameleon skin colour have been adapted by natural selection such that their proper functioning involves suppositions about aspects of their environment. If the earth had rotated faster, the circadian rhythms would have had to have been similarly faster, if the chameleon had been sitting in a different area, its skin would have had to have suitably different coloration.

In contrast, for a system to represent something there must be a complementary system to which the something is being represented. Similarly for a system to signal something there must exist a complementary system to which the something is signalled. These complementary systems must not be mere passive puppets, but must actively put these signals or representations to use within their own evolved agenda. In Millikanian terms, representation (or signalling) must involve both a producer, *and a consumer*, where each role has been selected for, rather than being merely a behavioural spandrel with respect to the signalling context.

The fourth relevant situation will be termed full-blooded signalling since it involves just such a pair of complementary systems (see figure 5.1d). Both members of the signalling dyad have been adapted by natural selection such that their functions involve the production and consumption of a signal which maps to some state of affairs. Perhaps S has evolved to give a twitch when holding a high-quality hand, *and* O has evolved to respond to S's twitches by folding her hand.[4]

What is at issue here is the *function* of the system within S which produces A under condition Q, and the *function* of the system within O which responds to A with R. Although information exchange, and fitness benefits are often associated with the evolutionary function of producing a signal, or consuming a signal, they are not definitional (i.e., necessary) characteristics of such systems, but merely *dispositional* characteristics. That is, information does *tend* to be exchanged during full-blooded signalling, and fitness benefits do *tend* to be accrued by signallers and observers within full-blooded signalling scenarios. However, there are instances of full-blooded signalling in which information is not exchanged, and similarly instances in which fitness benefits are not enjoyed by one, or either, member of a signalling dyad.[5]

The bee dance, performed in good faith, which misrepresents the location of a source of nectar (as a result of the fact that this nectar was consumed or moved during the dancing bees return flight) is one such situation. Neither information exchange, nor fitness benefits, are involved in this behaviour and the resulting response behaviours of the audience. However, the bee's misinformative dance is still a signal. This rather involved notion can be captured in the insight that signals sometimes *misrepresent* reality. The capacity to misrepresent is a defining characteristic of intentional entities. Signals, as prototypical intentional entities, must have this capacity. Millikan's thesis is that accounts which stress the informational content of signals leave no room for misrepresentation, and are therefore logically flawed.

Earlier it was mooted that without the notion that information exchange is central to signalling there was the possibility that hiding or camouflage might be classed as signalling, and that such a

---

[4] Such a state of affairs would appear to be unstable. See below for discussion of the issues which this observation raises.

[5] Compare — Hearts do *tend* to pump blood. However, there are instances in which a heart may fail to pump blood. Such a heart is still a heart, and as such its proper function is still to pump blood. It has simply failed to perform this function.

classification would be erroneous. However, an evolutionary-functional classification of signalling allows one to exclude such behaviours on the grounds that the mechanism which governs hiding behaviour, for example, does not function as a signal producer.

By the term function Millikan intends nothing more than the behaviour for which a system was selected (Millikan, 1989). If one selects a knife to open a bottle with, then the knife functions as a bottle-opener. Similarly, the function of my heart has been established by *natural* selection. Its function is to pump blood around my body since it was through pumping blood around my body that my heart's predecessors came to effect their reproduction and thus the current existence, and characteristics, of my heart.[6]

For Millikan, the function of a system can thus be ascertained by examining its adaptive history[7]. What kind of adaptive history must a system have in order to qualify for the function of signal producer, or signal consumer? A producer has been selected to behave in a manner which adapts a consumer to a state of affairs through generating a signal which maps with this state of affairs. It was through adapting the consuming system to this state of affairs that the ancestors of the producing system came to be reproduced and thus that the producing system came to be, and came to possess the constitution it does. Similarly, a consumer has been selected to behave in a manner which utilises the producer's signal as a means of adapting itself to the state of affairs represented by the signal.

These two functions are thus inter-twined; they implicate each other. Immediately a problem arises. How can such scenarios be achieved in situations in which there exists a conflict of interest between producer and consumer (S and O)?

Surely, given a conflict of interest between S and O, the selective pressures which drive signaller evolution are in opposition to those which drive observer evolution. This is the observation which led Dawkins and Krebs (1978) to suggest that, in such situations, both signaller and observer would evolve to exploit pre-adaptations in each other, and that such exploitation will continue as a co-evolutionary arms race.

Situations in which, unlike poker, each half of a signalling dyad has a co-operative interest in the other do not appear to involve this problematic instability. A classic example of such a communication system is the bee dance language. Since the reproductive potential of foraging bees completely coincide, there is no advantage to be gained in deceit. As a result, bee communication is stable, and can be interpreted under the scheme developed by Millikan. Dancing bees have evolved to produce a dance which maps onto the location of nectar in order to adapt the behaviour of their audience. Watching bees have evolved to adapt their behaviour on the basis of such dances, on the assumption that they map onto the location of nectar.

Thus, the signals passed between *co-operative* signallers appear to accord unproblematically with Millikan's definition of an intentional icon, and co-operative signalling scenarios may thus be

---

[6] N.b. The teleological flavour of this account is not damaging since such claims do not seek to account for the characteristics of my heart by claiming that they are the logical result of some unfolding process. Evolution does not strive to enable my heart to pump blood. However, that my ancestors' hearts have historically pumped blood, and that through this activity (rather than through being red, or making a bumpety-bump noise) my ancestors' hearts came to be reproduced, does have ramifications for functional descriptions of my heart (see Millikan, 1989, for a more developed account of evolutionary function).

[7] Obviously this cannot be done in any detail for any naturally evolved trait. However, the fact that there exists a particular adaptive history for any such naturally evolved trait implies that there is a corresponding particular (i.e., determinate) function for which the system has been selected.

counted as full-blooded signalling. However, situations in which there exists a potential conflict of interest between signaller and receiver do not seem to accord with such a definition. Mechanisms involved in the exploitation of sensory or motor biases (mind-reading mechanisms and manipulative mechanisms) do not appear to function as producers, or consumers of signals, under the definition of these functional roles provided above.

However, recall the definition of *conventional* signals offered in section 5.3. This definition cast conventional signalling systems as systems comprising two parts; a primary co-operative signalling system involving signallers and observers for whom there exist no conflict of interest, and a parasitic third party. Under the scheme provided by Millikan, such a characterisation of conventional signalling involves a full-blooded signalling component (those signals exchanged within the co-operative part of the system), and a non-signalling component (the behaviour of the parasitic third party) which relies on mere tacit suppositions concerning the character of the co-operative partners.

Handicap signalling systems may also be considered to involve full-blooded signalling. Within such systems, the character of extant signallers results from the success of ancestor signallers that possessed a system which produced signals which mapped with some state of affairs such that these signals were of use to some consuming system within observers (rather than producing signals which mapped that way for some adaptive reason external to the signalling context). These consuming systems found such signals useful (i.e., were selected to adapt the behaviour of observers to these signals in beneficial ways) because they mapped in the manner they did (rather than for some adaptive reason external to the signalling context). Although there is potential for a conflict of interest between signaller and receiver within handicap scenarios, the costs and benefits to both signaller and receiver result in a co-operative project being strategically sound for each partner. That is, for scenarios in which signalling is stabilised by costs and benefits, such that signallers are constrained to signal informatively, and observers are selected to take advantage of such information, it is true to claim that *honesty* is the best policy, rather than to maintain that such a signalling strategy is only honest *incidentally*.

Note on the Use of Examples and Mathematical Notation

As a cognitive scientist by training my biological intuitions are of dubious value. As a result I have frequently abstained from the use of examples of natural signalling except to occasionally note examples drawn from biological review articles.

However, I realise that the practice of matching theoretical results with empirical observations is extremely worthwhile. Indeed, within the biological signalling literature there are frequent statements to the effect that theoretical findings are waiting to be empirically tested, or that the construction of formal models of recent empirical findings is overdue. Within the theoretical literature there appears to be a degree of tension between those theorists for whom observation is the starting point for theory (e.g., Endler, 1991, 1993, perhaps), and those for whom the logical integrity of a theory is a first concern (e.g., Grafen, 1990a, 1990b, perhaps).

Throughout the following two chapters I will adopt the latter position. Although it is debatable whether natural examples of the systems which will be modelled exist, for the meantime I will concentrate on demonstrating the logical possibility of systems with the characteristics considered. The import of such demonstrations will be predicated on the fact that the models under examination have been published within the literature, and in at least one case (Grafen, 1990b), have become theoretical flagships within their field. However, it will be left to others, better qualified than myself, to debate the relevance of these findings.

Furthermore, all mathematical results will be derived within the body of the text rather than included separately in appendices. This strategy results from my belief that what mathematics is involved in the models is simple, and that postponing its presentation merely disrupts the development of any argument to which it is pertinent. However, this arrangement dose necessitate a degree of repetition, in that equations may need to be reiterated at some points. It is hoped that the reader will tolerate such repetition, and the degree to which proofs are presented as fully as possible.

Chapter 6

The Evolution of Symmetrical Signals

Recently, within the biology literature, there has been some interest in exploring the evolutionary function of animal displays through computer simulations of evolutionary processes. This chapter provides a critique of an exploration of the evolutionary function of complex symmetrical displays, investigating the hypothesis that complex symmetrical signal form is the product of a 'hidden preference' inherent in all sensory systems (i.e., a universal sensory bias). Through extending previous work and relaxing its assumptions it is revealed that the posited 'hidden preference' for complex symmetry is in reality a preference for homogeneity. The resulting implications for further accounts of the evolutionary function of complex symmetrical patterning are considered.

6.1 Introduction

A small but rapidly increasing number of studies (e.g., Arak & Enquist, 1993, 1995a; Enquist & Arak, 1993, 1994; Johnstone, 1994a; Krakauer & Johnstone, 1995; Hurd et al., 1995) have undertaken to explain evolutionary phenomena through the implementation of iterative computer programs, designed to model biological systems. These models differ from more traditional iterative numerical models in that they typically rely on modelling techniques pioneered within the artificial sciences (e.g., artificial neural networks, genetic algorithms, etc.). As such they may be considered as (proto-)evolutionary simulation models, although the authors of these models typically consider them to be little different from traditional iterative numerical models.

Some debate has accompanied this work (Cook, 1995; Johnstone, 1995; Arak & Enquist, 1995b; Dawkins & Guilford, 1995) but attention to the methodology employed within it has been cursory. This chapter replicates, discusses, and extends an evolutionary simulation model (Enquist & Arak, 1994) concerned with the evolutionary function of naturally occurring symmetrical visual displays (e.g., the complex symmetrical patterns found on the wings of many birds). First, a brief account of Enquist and Arak's (1994) argument advocating the interpretation of symmetrical signals as having evolved to exploit a universal sensory bias is presented. This is followed by a replication of their study and a discussion of its validity.

Subsequently an extension to their model will be introduced. This extension is designed to test the hypothesis that the networks evolved under the conditions employed by Enquist and Arak had preferences for structure (i.e., complex symmetry), only insofar as they were presented with patterns subjected to unnaturally structured sets of transformations during their evolution. The extended model supports this hypothesis, and demonstrates that a different account of Enquist and Arak's proposed sensory bias must be accepted. The chapter closes with some theoretical and methodological discussion. The roles of the sensory biases addressed within this chapter in the evolution of complex symmetry are considered, and some methodological issues concerning the application of evolutionary algorithms to biological modelling are raised.

6.2 Symmetry

Enquist and Arak (1994) attempt to show that symmetrical patterns are inherently favoured by perceptual mechanisms which must consistently classify objects despite viewing them from different angles, at different distances, in different orientations, and so on, due to the invariance of such patterns across such situations. Following Enquist and Arak, this postulated disposition to favour symmetrical patterns will sometimes be referred to as a (universal) sensory bias, in that it is proposed as an exploitable latent preference inherent within visual systems. However, this is not intended to imply that these visual systems' preference for symmetry need be established prior to, or in some way independently from, the associated signal evolution.

Enquist and Arak model an artificial visual system using an array of receptive elements (a seven-by-seven grid of idealised retinal cells each containing three receptors respectively sensitive to the red, green, and blue components of the light falling on their cell) and an artificial neural network consisting of 15 idealised neurons, each receiving 147 weighted inputs (one from each of the three receptors in each of the 49 retinal cells), which in turn excite or inhibit a decision neuron which responds 'yes' or 'no' according to whether the weighted activation from these 'hidden' neurons, when summed, and perturbed slightly by a random internal factor, rises above a certain threshold.

This visual system's task was to accurately discriminate between a signal pattern and various distractor patterns (each represented as a five-by-five array of coloured pixels; each colour being represented by a vector in the red-green-blue colour space) despite each being presented to the system in various positions on the system's 'retina' and in various orientations. These positions and orientations were arrived at through subjecting the signal (and the distractors) to reflection, rotation, and translation transformations (see figure 6.1).

Artificial selection pressures were applied to such visual systems and the signals they attempted to classify (both of which were initially randomly configured). Repeatedly, the best of each (i.e., the network that discriminates signal from distractor under the most transformations and the signal that is discriminated correctly under most transformations) were allowed to 'asexually reproduce', creating new patterns and networks that were each similar to their parent but not identical due to low probability 'mutation' events which arbitrarily corrupted the reproduction process.

Under this artificial evolutionary paradigm impressive results were obtained. Patterns and networks co-evolved over evolutionary time, reaching "... a quasi-stable situation in which the network discriminates almost perfectly ... and the signal itself changes slowly ... [The signals] consisted of purer, brighter colours than random patterns; they also displayed ... marked symmetries" (Enquist & Arak, 1994, p.171).

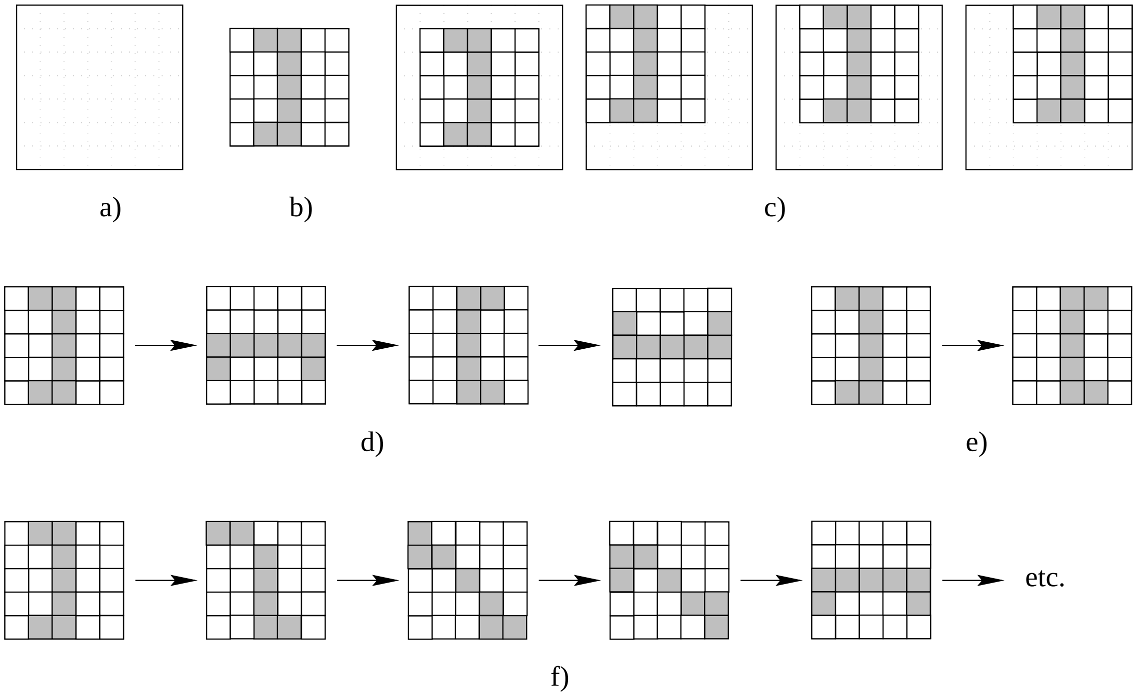


Figure 6.1: Depicting (*a*) the retina, (*b*) an example signal, Enquist and Arak's (1994) (*c*) translation, (*d*) 90° rotation, and (*e*) reflection transformations, and (*f*) the higher resolution rotation transformation employed in this chapter (n.b. in (*c*) four of the nine possible translations are depicted, and in (*f*) five of the 16 orientations are depicted).

Prima facie these results are strong support for the hypothesis that symmetry "may arise as a by-product of the need to recognise objects irrespective of their position and orientation in the visual field" (Enquist & Arak, 1994, p.169).

6.3 Replication

In replicating this study the stochastic hill-climbing algorithm employed by Enquist and Arak (1994) was re-implemented[1]. The design of the simulation presented here should thus be attributed to Arak and Enquist. Discussion of the adequacy of this design will be deferred until section 6.7.

Initially a random network (with connection weights drawn from a uniform distribution $[-0.3, 0.3]$) and a random signal (colour components drawn from a uniform distribution $[0.0, 1.0]$) are generated. The algorithm then repeatedly carries out two consecutive periods of hill-climbing, one within the space of signal patterns and one within the space of retinal networks.

Hill-climbing is carried out in the following manner. A mutant of the current best network is generated by copying each of the best retina's network weights with probability 0.01 of a weight being perturbed by a normally distributed value (zero mean, s.d. 0.02). The mutant is assessed with respect to the current best signal, and, if discovered to be fitter than the current best network, takes the best network's place, acting as the model for the next mutant. Once *n* mutants have been generated and tested in this manner, attention is switched to the best signal pattern.

---

[1] Copies of the code, and colour versions of this chapter's figures, are available from the world-wide-web page `http://www.cogs.susx.ac.uk/users/sethb/sym.html`

The signal is tested with respect to the current best network and compared to a mutant which is generated by copying each colour component of the signal with probability 0.05 of perturbing it with a normally distributed value (zero mean, s.d. 0.3). Perturbations which produced colour component values outside the legal range $[0.0, 1.0]$ were truncated to the nearest extreme value[2].

If a mutant signal is assessed as fitter than the current best signal, it takes the best signal's place and acts as the model for the next mutant. After n mutants have been generated attention reverts to the best retinal network and the process repeats.

Although the retinal network evolution is described first here, in order to avoid biasing our replication through preferentially evolving either the network or the signal first, the order in which the evolutionary 'steps' were carried out was counter-balanced across simulations[3].

Enquist and Arak dub one cycle of the above process (in which, arbitrarily, n=100) a 'generation', and run their simulations for 500 such generations. The simulations described here employed fewer iterations within a generation, and fewer generations (n=50 for 100 generations) after pilot studies revealed that when simulations were run for 500 (n=100) generations, there was no significant change in the makeup of either the networks or the signals after the first 50 such generations.

Fitness was calculated thus,

$$f(s,r) = \frac{1}{p}\sum_{i=1}^{p}\left[\xi(\phi_r(\lambda_i(s))) - \frac{1}{q}\sum_{j=1}^{q}\xi(\phi_r(\lambda_i(d_j)))\right]$$

where $f(s,r)$ denotes the fitness score ascribed to signal and retina $\{s,r\}$ under an assessment regime comprising p presentations of different transformations of both signal s and q distractor patterns (d_j denotes the j^{th} distractor pattern). The compound function $\phi_r(\lambda_i(x))$ denotes the output of retinal network r when presented with pattern x under transformation λ_i.

This output is calculated as follows. The output of a retinal receptor was equal to the intensity of its input, whilst that of a network neuron was calculated as a sigmoid function of y, the sum of its weighted inputs, thus:

$$\text{output} = \begin{cases} 1 - 1/2[1/(1+y)] & y \geq 0 \\ 1/2[1/(1-y)] & y < 0 \end{cases}$$

The function $\xi(w)$ is an error function which smoothes the network output (w) by modelling a stochastic internal variable (Θ) which perturbs the network output by a value normally distributed with zero mean (μ=0), and standard deviation 0.02 (σ=0.02).

Rather than draw a random value from this distribution for each trial, the effect of this internal factor was calculated in the limit, thus,

$$\xi(w) = P(w + \Theta \geq \theta) = \frac{1}{\tau}\int_{\theta}^{\infty} e^{-\left(\frac{z-w}{\sigma\sqrt{2}}\right)^2} dz$$

where τ is a scaling constant used to normalise the function's output to between 0 and 1, whilst θ is a threshold value above which the network is said to have responded positively to its input.

---

[2] The details of this mutation operator (which was employed throughout all the simulations described here) were not reported by Enquist and Arak (1994), who in addition to employing this scheme, entertained a number of alternative mutation operators (also unreported, Enquist & Arak, pers. comm.). See section 6.7.1 for discussion of the implications of these mutation operators.

[3] Enquist and Arak (1994) do not report whether they employed this counter-balancing measure, or, if not, which population was allowed to hill-climb first. See section 6.7.

The value returned by the error function was thus a deterministic measure of the proportion of trials in which a network output value (w), after perturbation by an internal stochastic variable (Θ), would exceed the arbitrarily designated threshold value (θ).

For all simulation runs considered here, $\theta=0.5$, $q=3$ (one white, one black, and one random distractor)[4].

Enquist and Arak's (1994) presentation regime was replicated. Each pattern was presented to a network under 36 different transformations (nine translations by four orientations, see figures 6.1c and 6.1d). The results of 20 simulation runs are displayed in figure 6.2a.

Bilateral asymmetry was calculated as the average geometric distance in the red-green-blue colour space between corresponding pixels. Rotational asymmetry was calculated as the mean of the standard deviations of the groups of pixels invariant under the rotation transformation. Brightness or boldness was calculated as the average geometric distance in the red-green-blue colour space between each pixel and the nearest corner of the colour space.

Evolved signals exhibited significantly more bilateral symmetry ($t=6.526$, d.f.=19, $p<0.01$), four-fold rotational symmetry ($t=9.638$, d.f.=19, $p<0.01$), and brightness ($t=10.825$, d.f.=19, $p<0.01$), than random signals (see figure 6.3a). Under the conditions employed by Enquist and Arak (1994), symmetrical signals do indeed evolve.

6.4 Discussion

Objects are not always seen at the same orientation, or from the same distance, etc. The transformations that Enquist and Arak's (1994) patterns undergo are intended to mimic these different ways in which the light reflected from a real-world object falls onto a retina. As Enquist and Arak's retinae and patterns are square, reflections in the vertical and horizontal axes seem natural transformations to perform in order to create variation in the retinal image. However, natural visual systems are very rarely exposed to a pattern and then, subsequently, a reflection of that same pattern.

Rotations of 90 degrees also seem natural transformations to perform upon square patterns. In response to such rotations, Enquist and Arak's simulation produces patterns with four-fold rotational symmetry. However, in reality, the rotation transformations that images undergo are *continuously* (although not necessarily uniformly) distributed between 0° and 360°.

Finally, if a signal's invariance under transformations is what makes it easy to discriminate, then biological signals which are selected for conspicuousness should ideally be one solid colour (i.e., maximally symmetrical).

To summarise, could it be the case that the networks evolved under the conditions employed by Enquist and Arak had preferences for structure (i.e., complex symmetry), only insofar as they were presented with patterns subjected to unnaturally structured sets of transformations during their evolution?

---

[4] The exact number and nature of the distractor patterns employed by Enquist and Arak (1994) is unclear, and varied over the simulations they report. See section 6.7.

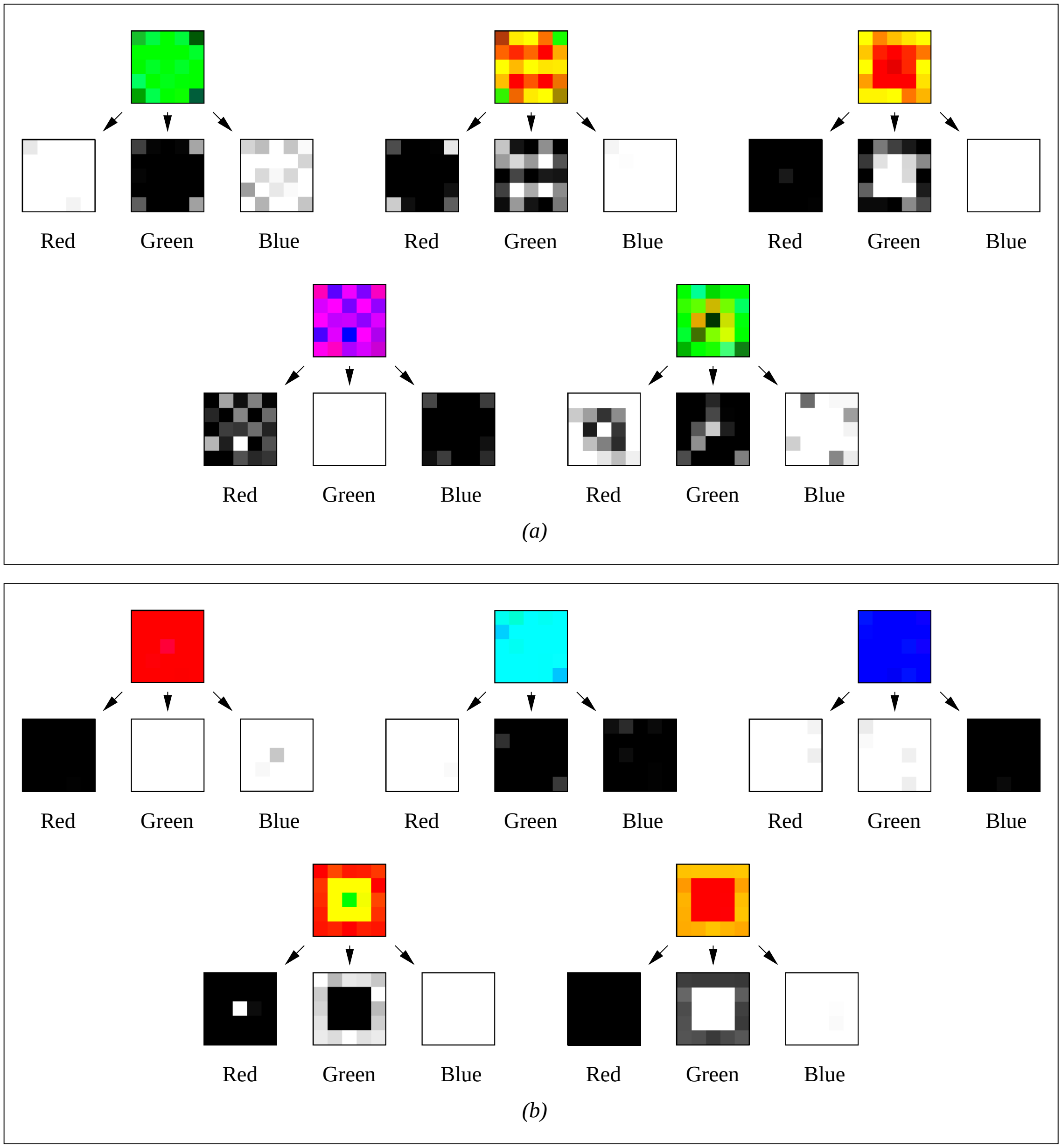


Figure 6.2: Five signals typical of (*a*) the 20 signals evolved under Enquist and Arak's (1994) presentation regime, and (*b*) the 20 signals evolved under a presentation regime comprising nine translation, and sixteen 22.5° rotation transformations. Each signal is shown as a colour composite and three colour separates in which the increasing intensity of each colour component is represented as increasingly heavy shading.

6.5 Extension

In order to explore the effect of more continuous rotational transformations, the rotation regime was relaxed to include transformations that fell at 22.5° intervals (see figure 6.1f) rather than the 90° intervals employed by Enquist and Arak (1994). This quadrupled the number of presentation trials per assessment (i.e., p=144), as, under each translation transformation, each signal was now presented in sixteen, rather than four, different orientations. All other parameter values were held constant. The results of 20 simulation runs are displayed in figures 6.2b and 6.3b.

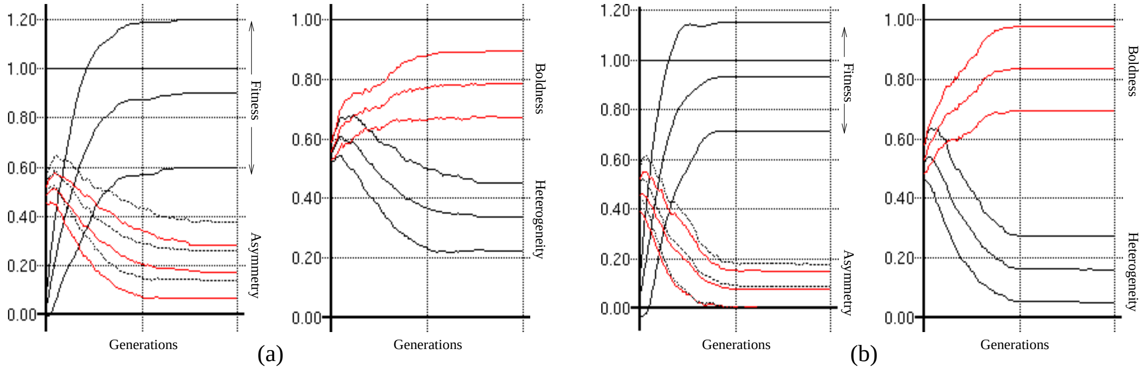


Figure 6.3: Initially random signals were subjected to (*a*) Enquist and Arak's (1994) presentation regime comprising nine translation, and four 90° rotation transformations, and (*b*) a presentation regime comprising nine translation, and sixteen 22.5° rotation transformations. Graphs depict (left) plots (+/- s.d.) of signal fitness, and rotational asymmetry metrics (four-fold and sixteen-fold) averaged over 20 simulations of 100 (*n*=50) generations each, and (right) plots (+/- s.d.) of signal boldness and heterogeneity metrics for the same 20 simulation runs.

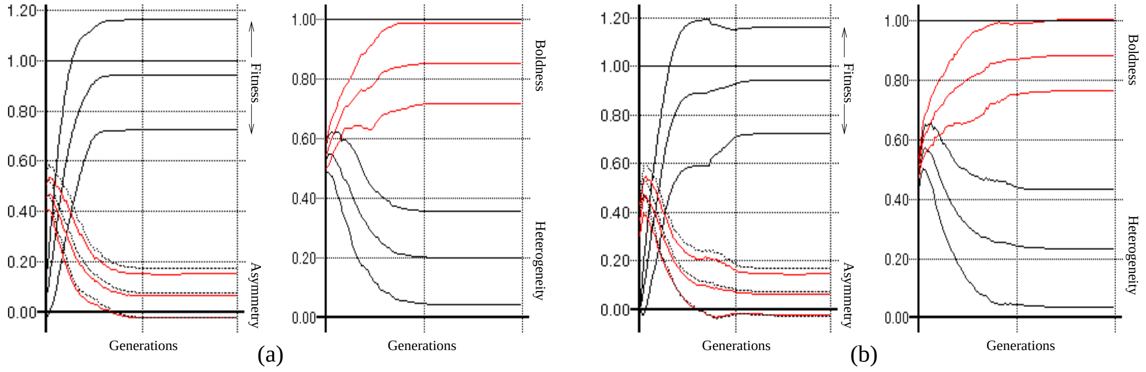


Figure 6.4: Initially (*a*) random, and (*b*) bilaterally symmetrical signals were subjected to a presentation regime comprising nine translation, and sixteen 22.5° rotation transformations. In addition, signals depicted in graphs (*a*) were subjected to a bilateral reflection transformation. Graphs depict metrics as per figure 6.3.

Evolved signals display significantly greater rotational symmetry, both sixteen-fold (t=5.07, d.f.=19, p<0.01), and four-fold (t=3.3, d.f.=19, p<0.01), and are significantly more homogeneous (t=4.87, d.f.=19, p<0.01), than those evolved under the conditions employed by Enquist and Arak. They also feature significantly brighter colours than random signals (t=10.135, d.f.=19, p<0.01). In a further 20 simulation runs, introducing a bilateral reflection transformation (see figure 6.1e), which doubled the number of presentation trials per assessment (i.e., *p*=288), did not result in evolved signals significantly different in any way from those generated under this sixteen-fold rotational regime (see figure 6.4a).

Furthermore, seeding 20 simulations with initially perfectly bilaterally symmetrical signals also resulted in evolved signals not significantly different from those generated from random initial conditions, indicating that, as well as failing to *generate* complex symmetrical structure, the

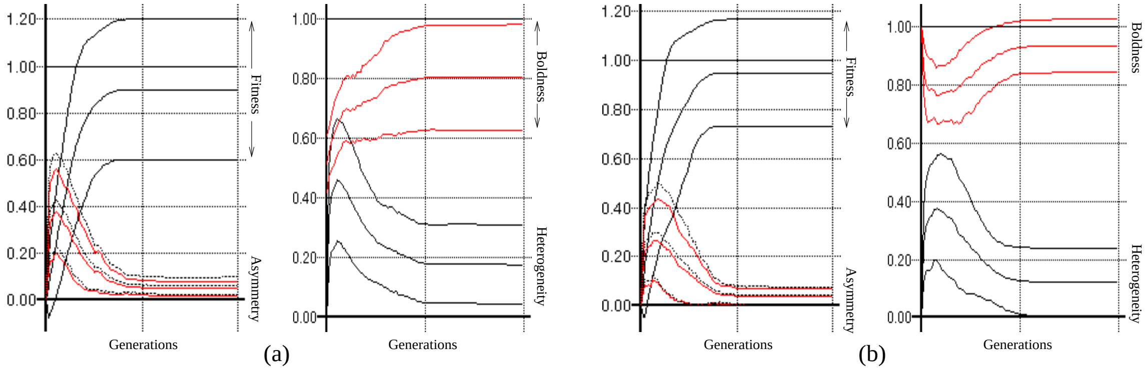


Figure 6.5: Initially (*a*) homogeneous signals and (*b*) maximally bold homogeneous signals were subjected to a presentation regime comprising nine translation, and sixteen 22.5° rotation transformations. Graphs depict metrics as per figure 6.3.

sensory bias presented by Enquist and Arak (1994) could not *maintain* pre-existing bilateral symmetries (see figure 6.4b).

Further experimentation reveals that not any homogeneous signal suits any network. Seeding 20 simulations initially with signals homogeneous in a random colour resulted in evolved signals not significantly different from those evolved from random initial signals under the same presentation regime. However, as networks evolve to prefer signals of above average brightness, the final evolved signals were significantly more bold than their earliest ancestors (t=10.734, d.f.=19, $p<0.01$; see figure 6.5a).

Seeding 20 simulations with signals initially homogeneous in one colour chosen randomly from the six maximally bold colours that the networks were not penalised for favouring (i.e., one of green, red, blue, magenta, cyan, or yellow, but not white or black) again resulted in evolved signals not significantly different from those evolved from initially random signals (see figure 6.5b). However, the corner of the colour space occupied by the initial signal was not necessarily the same as that occupied by the final signal. Thus, the networks were not merely favouring any homogeneous, bold signal, but initially had innate biases which could result in signals evolving from one bold, homogeneous pattern (e.g., solid green) to another (e.g., solid blue).

That this is the case is represented graphically in figure 6.5b by (i) the increase in asymmetry and heterogeneity (which are initially zero, since ancestral signals are maximally homogeneous), and the associated decrease in boldness (which is initially maximum, since signals are comprised of extreme colour components) which occurs at the *outset* of the simulation, as signals evolve away from their maximally bold, maximally homogeneous ancestors, followed by (ii) a drop in asymmetry and heterogeneity, and a rise in boldness, as signals converge upon homogeneous, bold patterns which may differ from their ancestral seed.

In summary, signal form was shown to be a result of pressure, not for symmetry *per se*, but for homogeneity coupled with arbitrary boldness (the nature of which being determined by the initial constitution of the network), which *coincidentally* results in signals which exhibit a high degree of both rotational and bilateral symmetry. As such, complex signal form cannot be explained as a product of the sensory bias that Enquist and Arak (1994) propose.

6.6 Theoretical Discussion

Several points should be made about the scope and nature of this result. This section will address some of the theoretical issues raised by the failure to replicate Enquist and Arak's (1994) results. The subsequent section will discuss some methodological points pertaining to the proper construction and execution of evolutionary simulation models.

First, this result should not be taken to be an explanation for the presence of homogeneous patterning in the natural world. Although certain cases of bold, homogeneous coloration may well be adaptations exploiting the conspicuousness of such patterns, no general claim to that effect is being made here. The thesis of this chapter is purely negative; complex, symmetrical patterning cannot be explained as the result of evolutionary pressure to exploit the universal sensory bias proposed by Enquist and Arak (1994).

Secondly, it might be thought that a presentation regime in which the orientation of signals was biased in some way could lead to more complex symmetry. Consider a presentation regime in which the orientation of the signal is drawn from a normal distribution centered on some arbitrary angle (see figure 6.6a). This can be thought of as modelling a situation in which the signal has a typical orientation, but sometimes deviates from it slightly. Under such a regime, might bilaterally symmetrical signals evolve?

Enquist and Arak's model ensures that in order that one part of a signal comes to resemble some other part (i.e., that some symmetry develops), both parts of the signal must at different times fulfil the same role for the signal receiver. That is, both signal parts must be projected onto the same part of the receiving retina.

If we choose a candidate axis of symmetry passing through the centre of a signal, and lying within the plane occupied by the signal, we see that whilst light reflected from signal areas on either side of this axis might sometimes fall on the same part of a receiver's retina, the chance of this occurring decreases as the angular distance between such signal areas increases. For example, a presentation regime in which signal orientation is drawn from some normal distribution centred upon some typical orientation will ensure that light reflected from signal elements at $\pm 5^\circ$ from an arbitrarily chosen axis will fall on the same area of retina far more frequently than light reflected from signal elements at $\pm 45^\circ$ from the same axis. As a result, signals will gain only very limited benefits from any degree of bilateral symmetry. Since this is true for any axis passing through the centre of the signal, no one axis of symmetry has precedence over others, and reflectional or rotational symmetry will only be favoured in that it increases homogeneity (homogeneous signals still being easier to discriminate from distractors than more complex signals).

However, consider a probability distribution function governing signal orientation which is bimodal, with the angular distance between modes equal to 180° (see 6.6b). Such a distribution, whilst again not affording precedence to one particular axis of reflectional symmetry, will induce two-fold rotational symmetry (although not at the expense of homogeneity) as signals which are invariant under a 180° rotation will be favoured. The orientation aspect of Enquist and Arak's presentation regime can be considered to be a probability distribution function with four equidistant modes. Enquist and Arak (1994) demonstrated that such a presentation regime resulted in signals exhibiting four-fold rotational symmetry.

However, it cannot be claimed that such multi-modal probability functions underly many nat-

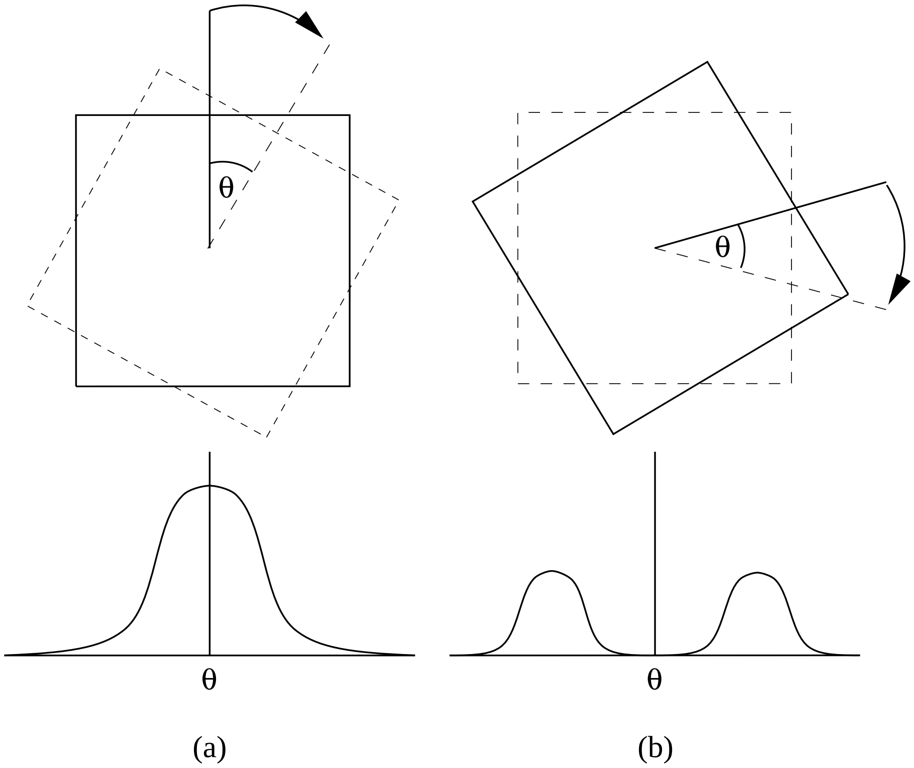


Figure 6.6: Depicting probability distributions governing the orientation of a display signal. Distribution (*a*) is a unimodal normal distribution, whilst distribution (*b*) is a bi-modal distribution. Although a reference orientation for the signal and an associated reference line (both depicted with solid lines) have been chosen for each diagram this orientation is arbitrary, e.g., for (*a*) $\theta = 0$ denotes an upright signal and θ is measured as the angular displacement from a vertical reference line, whereas for (*b*) $\theta = 0$ denotes a signal 60° off vertical and θ is measured as the angular displacement from a line 15° from horizontal. For any distribution such as (*a*) or (*b*), once a signal orientation has been chosen for $\theta = 0$, theta may be measured from any any reference line without affecting the presentation regime of the signal. Such reference lines cannot be considered as 'axes' with respect to the rotation operator, since the only axis of rotation is perpendicular to the plane of the signal.

ural signalling scenarios. For example, consider a patterned fish swimming from right to left across a potential mate's visual field. In order to present the same pattern at an angular offset of 180°, the fish would have to swim upside down (i.e., rotated through 180°). For the fish to provide a four-mode display similar to that employed by Enquist and Arak, it would have to spend time swimming vertically up and vertically down as well as right-side up and up-side down (or some other combination of four equi-distant orientations), whilst spending comparatively little time swimming at the intervening orientations. Note that a fish merely swimming from right to left and subsequently left to right would not display a pattern at two orientations, nor a pattern and subsequently a reflection of that pattern, but simply two patterns (one from each side of the fish). Whilst the patterns on each side of such a fish might be related, such a relationship is contingent rather than necessary, and is not modelled by Enquist and Arak's scheme which does not include morphological/developmental constraints on pattern form (see below).

It is not impossible to imagine the discovery of naturally occurring signalling behaviour with a complex stereotypical structure (e.g., an animal that displays a patterned body part at four distinct orientations) coupled with an associated discovery that the structure of this signalling behaviour is reflected in the structure of the displayed signal patterns. However, the complex structure of such signals could not be said to have arisen "as a by-product of the need to recognise objects irrespective of their position and orientation in the visual field" (ibid., p.169). On the contrary, the complex structure of such signals would have arisen *despite* such a 'need'. In such a case, the general pressure which favours homogeneity as a response to this 'need' would have been relieved through restricting the presentation of the signal in some specific (and unusual) manner. Such complex signal patterns could not be explained as the consequence of "general principles of perception" (ibid., p.169), but could only be explained through reference to specific facts about their evolutionary history.

Finally, a further consideration which might influence the expression of a bias for homogeneity is the degree to which the developmental or morphological constraints of a creature might interfere with the form of its signals. It is attractive to argue that a creature's bilaterally symmetrical patterning might be the result of constraints imposed by its bilateral *Bauplan*. Could it be that a sensory bias for homogeneity might be expressed across a bilaterally symmetrical body as similarly bilaterally symmetrical patterning?

Augmenting Arak and Enquist's account of complex patterning with an appeal to morphological constraints is unsatisfactory as such an appeal must posit constraints which *resist* homogeneity. For example, if morphological constraints which impose bilaterality upon a creature's morphology can be shown to influence the creature's patterning such that the patterning is similarly bilaterally symmetrical, then any 'hidden preference' for homogeneity may be regarded as superfluous to an explanation of such complex bilateral symmetry. An adequate explanation may be based entirely on the aforementioned morphological constraints. Furthermore, a sensory bias for homogeneity will actually resist the development of complex bilaterally symmetrical patterning in favour of more homogeneous bilateral patterns (i.e., uniform coloration).

6.7 Methodological Discussion

If models are to provide theoretically worthwhile results, the methodology employed in their design and execution must be rigorous enough to withstand critical accusations of artefactual results due to poor simplifying assumptions. Although simplifying assumptions are a necessary part of modelling (whether through simulation or formal analytic methods), they must be explicit and well justified. Moreover any conclusions drawn from such models must be qualified with respect to those simplifying assumptions.

The assumptions made by Enquist and Arak (1994) involving their presentation regime, their mutation operator, their representation of colour, their implementation of noise, and their model evolutive process, all fail to survive close scrutiny, and in drawing conclusions from their results the authors fail to appreciate the considerable part played by these assumptions.

It has been shown above that the form of the signals that Enquist and Arak (1994) evolve is dependent on the presentation regime that they undergo. The signals exploited the structure inherent in the *presentation regime* rather than any 'hidden preferences' inherent in their artificial retinae. In this section the effects of their additional simplifying assumptions are discussed.

6.7.1 Mutation

The term mutation is used within biology as a label for a diverse range of deleterious processes which interfere with the veridical transcription of genotypes. Cosmic rays, transcription errors, toxins, etc. are all implicated in the creation of mutants. Under Darwinian rather than Lamarckian schemes, such mutants are assumed to differ from their parents in a non-directional manner, rather than being biased in some direction of increasing fitness.

Evolutionary simulations involving asexually reproducing populations typically rely on mutation operators to provide genetic diversity upon which selection may act. Simulations involving sexual populations similarly involve mutation operators, but augment them with crossover operators which allow offspring to inherit passages of genome from each parent.

Although the construction of mutation operators is not typically regarded as a complex matter, such operators are often significantly different from natural mutation processes, and involve issues which are unique to the design of evolutionary simulations.

Boundedness and Legality

Consider a phenotypic trait which may vary over some range, and is represented by some real value in an artificial genotype. (Alternatively, traits may of course be represented with a fixed or variable number of binary, (or decimal, or octal, etc.) digits. Our attention will be confined to real-valued genes initially since they were employed by Enquist and Arak, and will also be employed within the models presented in the next chapter. Subsequently the discussion will be extended to address discrete genetic representations of phenotypic traits).

We may distinguish between two classes of such a trait. Unbounded traits are those which may vary limitlessly. Such traits might include, perhaps, the significance afforded to a male display by a female onlooker, which might vary from zero (non-significant), through positive infinity (infinitely attractive), or negative infinity (infinitely repellent).

Bounded traits are those for which there are a limited number of legal values. Many traits suffer a lower limit. For example, one cannot have less than zero legs. Some are bounded at both extremes, e.g., a trait governing the time of day at which one begins to forage might only vary between dawn and dusk, or, as under Enquist and Arak's (1994) scheme, the colour component of a signal element might only vary between zero and unity. Such a notion of boundedness raises the associated notion of legal and illegal genotypes, illegal genotypes being those that code for traits which lie outside their legal bounds.

It is clear that the distinction being made here is an unnatural one which cannot easily be applied to natural genetic encodings. For example, although natural genes may code for what appears to be a bounded phenotypic trait (e.g., the length of a tail feather), the manner in which they do so may logically preclude the occurrence of illegal values for this trait. Additive polygenic traits, for example, might code for a continuously varying phenotypic trait with a varying number of particular genes. Since the genotype cannot contain a negative number of these genes, it cannot code for an illegal phenotypic trait (e.g., a tail of negative length).

In one sense however, an 'illegal genotype' is one which does not result in a viable organism.

Any mutant genotype will be subject to morphogenetic processes. Some will fail to be viable. Such mutants will not leave offspring. As a result they will be selected out of the population. This selective process is essentially the same selective process which excludes viable mutants which are less well adapted to their niche than their competitors.

Genotype legality within artificial evolutionary algorithms falls somewhere between the accounts given in the two previous paragraphs. It is true that, as in the first account, genotypes with illegal parameter values are never realised as organisms, and are thus never subjected to the same selective pressures as their valid conspecifics. However, illegal genotypes are generated by the algorithm and are thus not excluded out of logical necessity as in the first account, but are selected as invalid, as in the second account. However, the grounds upon which illegal genotypes are selected are not the same as those which govern phenotypic selection. The legality of genotypes is assessed, prior to any morphogenetic, developmental or ontogenetic performance, on the basis of the genotype itself. If found to be legal, the genotype is translated into a phenotype and assessed as normal. These two selection procedures are separate unlike the single selective process described above. If found to be illegal, some alternative course of action is taken (see below).

It is this lack of correspondence between artificial evolutionary algorithms and natural evolution which leads to the the unnatural character of artificial mutation operators and the associated selection biases which they introduce.

Mutation Biases

The designer of a mutation operator has several decisions to make concerning its character. First, what is the effect of a mutation event upon a parameter value? Should a mutant value be drawn randomly from the range of legal values in a manner independent from its pre-mutation (parental) value? Or should the mutant value be related to this parental value, or some other value (e.g., the population mean), and if so, how?

The answers to these questions are influenced by prior decisions concerning the underlying genetic representation, and the boundedness of the phenotypic traits which the genes represent.

Real-valued genes coding for bounded phenotypic traits may be mutated independently of the genetic makeup of the current population through drawing a random value from the range of legal values available for a trait.

Similarly, real-valued genes coding for unbounded, or partially bounded, phenotypic traits can be mutated independently of their parental value through imposing some arbitrary legal range upon the phenotypic trait and picking a value from this range.

However, the appeal of mutation operators which generate mutant values independent of their parental value, or the population mean, or the general makeup of the population seems limited. Often, designers of mutation operators wish mutated phenotypic traits to differ only slightly from the parental phenotypic trait. This motivation appears to underlie the adoption of perturbation style mutation operators for real-valued genes, and Gray coding for binary coded traits.

Traits encoded as groups of binary digits which are interpreted as phenotypic values (e.g., two bits interpreted as between zero and four eyes), are typically mutated via random bit flips. For such mutation operators acting upon conventional binary numbers, mutated phenotypic traits are poorly correlated with parental phenotypic traits, since a single bit-flip may result in a large change in the value for which the bit-string codes. Gray coding (Gray, 1953), through ensuring that consecutive integers are coded for by adjacent binary strings, increases the correlation between mutants and their parents. This coding scheme is advocated by genetic algorithm designers, who claim it makes evolutionary search more effective (Goldberg, 1989).

Mutation operators (for both real, and binary, encoding schemes) which generate mutant values which are influenced by parental values rather than legal limits have the potential to generate illegal parameter values for bounded traits. For such mutation operators, decisions concerning the treatment of such illegal mutations must be made. There are a number of options that an operator designer might take.

The first is to ignore mutation events which generate illegal values and leave the value unchanged. The second is to generate as many mutant values as is necessary until a legal value is arrived at. The third option is to truncate any illegal mutant value to the nearest legal limit. The fourth, and most natural option is to reject any offspring which include illegal parameter values, and generate new ones from scratch, re-choosing parents.

Various of the first three options were employed by Enquist and Arak in constructing a mutation operator for the colour components of the evolving signals. Legal values for the colour components lay within the range $[0.0, 1.0]$. Mutant values were generated with probability 0.05 by perturbing the parent component with a normally distributed value (zero mean, s.d. 0.3). The third option was employed exclusively throughout the simulations presented here.

Each of these options has unwanted side-effects which may influence the evolutionary dynamics of a simulation. These side-effects share a common character. Each results in edge effects at the extremities of a parameter's legal range.

Option one operators ignore mutation events which generate illegal values. For any parameter, more illegal values will be generated for parent values near the extremes of the legal range. This ensures that the effective mutation rate (i.e., the rate at which offspring inherit mutated parameter values) will decrease as parental values approach legal extremes.

Conversely, option two operators, which generate as many mutant values as is necessary until one is valid, will maintain a constant effective mutation rate across the spectrum of parameter values, but will tend to mutate extreme parental values away from the legal limits of the parameter.

Option three operators will generate an effective mutation rate which is constant across the range of parameter values, save that it falls to half the nominal mutation rate at the extremes of the legal parameter range. Parameter values near the extreme will tend to reach the legal boundary for the trait (unlike the previous two options for which it is very unlikely that the extreme values will be reached at all) and be kept there.

Option four operators, although seemingly the most accurate reflection of natural mutation processes, result in a selection pressure which resists the evolution of extreme parameter values. This can be seen by considering the frequency of viable offspring which parents can expect to generate. As parental values approach an extreme, the frequency of non-viable offspring increases, reducing their fitness. Thus, all other things being equal, extreme-valued parents will more often leave no offspring than their more conservative-valued competitors.

These worries may seem rather overworked. However, they reveal the potential for mutation biases to influence the results of evolutionary simulations. For example, the "purer, brighter colours" (Enquist & Arak, 1994, p.169) of signals evolved under Enquist and Arak's presentation regimes (i.e., the extreme values of the evolved colour components) might well be explained as an artefact of the mutation operator which they employed. The colour components were represented by bounded real values in the range $[0.0, 1.0]$, and subjected to fairly large mutation events (perturbations normally distributed with standard deviation equal to 0.3 ensure that roughly a third of mutation events would perturb the parental value by more than 0.6, i.e., over half the maximum range for legal values). This, coupled with a mutation operator which lowers the effective mutation rate as parental values approach the legal boundaries of a parameter, could lead to signals with "purer, brighter colours" merely through genetic drift.

The presence of potential artefacts such as this can be detected through exploring the effects of differing initial conditions. In order to rule out the account of bold coloration as an artefact brought about by mutation bias, simulations were run from initial conditions involving maximally bold, homogeneous signals (see figure 6.5b). These simulations demonstrated that signals were able to evolve away from extreme-valued ancestors in response to co-evolutionary selective pressures provided by arbitrary network preferences.

6.7.2 Colour

Enquist and Arak (1994) mention the bright colours of their evolved signals but wisely provide no adaptive explanation for them. The fact that colours in their simulation are represented as vectors in a simple three-dimensional colour space, and evolve by means of large mutations, typically along a single dimension of this space, render any conclusions drawn from their results at best tenuous. Having ruled out mutation bias as an explanation, the colours present in the evolved signals are best thought of as resulting from predispositions of simple artificial neural networks to favour extreme-valued inputs, a characteristic of artificial neural networks which may not generalise to natural neural architectures (Dawkins & Guilford, 1995).

6.7.3 Noise

Simulation design considerations are subject to two often conflicting influences. Simulations must give evolution free rein to explore genotype space, but not at the expense of biological realism. On the one hand simulations must not involve constraints which obstruct adaptive change. The genetic code, mutation operator, etc., must not introduce biases or constraining factors which might interfere with evolutionary dynamics. On the other hand, simulations must not involve processes which preclude the effects of valid retarding evolutionary forces. Crudely put, simulations must neither artificially impede evolution nor artificially assist evolution.

The random internal factor implemented by Enquist and Arak (1994) serves to smooth the fitness landscape and enables evolutionary adaptation over areas of the landscape which have a very low fitness gradient. This improvement in evolutionary search is achieved through calculating the effect of noise in the limit.

By contrast, modelling the noise as a random value drawn independently from some distribution each time the internal stochastic variable is invoked, results in evolved signals only slightly more symmetrical than random ones, i.e., the effect of the selection bias favouring symmetrical/homogeneous signals appears to be much reduced.

Under this condition, evolution towards discriminable signals and discriminatory networks is impeded because slight improvements in discriminatory ability on the part of the retinal networks or discriminability on the part of signal patterns are not consistently scored as fitter than marginally poorer competitors.

For example, a mutant network which, before the effects of noise are considered, classifies a signal with an output (e.g., 0.48) which falls below the output threshold (e.g., 0.5), but which is slightly higher than its parents output for the same signal under the same presentation conditions (e.g., 0.46), might sometimes (rarely) enjoy a random perturbation (e.g., +0.03) which lifts the output above the threshold and changes the network's response from incorrect to correct. Since the mutant's output is closer to the threshold than that of its parent's, it will classify this particular signal correctly (slightly) more often than its parent.

If such perturbations are drawn randomly for each trial, next time the network is exposed to the very same signal under the very same conditions, it will generate a different output value (e.g., 0.48+0.01), which may lie below the output threshold and will thus not result in a correct classification. Such a scheme ensures that networks will be assessed with fitness values which, although non-deterministic (i.e., involving a random element) reflect the ability of the network to make a binary decision task in the face of some internal variability (noise on the output node). The fact that noise values are typically very small ensures that an increase in fitness resulting from the beneficial effects of noise will be vanishingly infrequent for any output values not extremely close to the threshold before the addition of noise.

Such a scheme contrasts starkly with Enquist and Arak's deterministic model which rewards *any* improvement in performance with an improvement in fitness on the basis that over an infinite number of trials any improvement, however small, *must* result in more correct discriminations.

Over evolutionary time, a system under directional selection for some trait will benefit from the slightest tendency to increase such a trait, despite such a tendency resulting in an increase (and thus improved performance) only very occasionally. Such tendencies, even if they are not consistent, will spread through the population if, on the average, they provide some improved performance. It is this reasoning which underlies Enquist and Arak's adoption of an internal stochastic variable modelled in the limit.

However, in a co-evolutionary situation, the spread of a trait which allows an improvement in some behaviour only occasionally is necessarily influenced by changes in the co-evolving population, which will not necessarily be stable over the number of generations necessary for the fixation of such traits.

For the scenario under examination here, an increased tendency to discriminate signals from distractors which only results in an increase in correct judgements once in every 10 generations may never fixate in the discriminatory population since the signals which the network is better adapted to discriminate may themselves have changed before the network's worth can be established by natural selection.

Thus, within this model, the addition of noise effectively changes the nature of the discrimination task from that of a binary decision task (signal vs. distractor), which rewards networks with discrete units of fitness which are large with respect to typical fitness scores, to that of a continuous judgement task, which rewards networks with smoothly varying fitness increments. The former is obviously a more natural candidate for modelling the natural scenario which Enquist and Arak describe, whereas the latter affords more effortless evolutionary search.

6.7.4 Search

Issues which arise as a result of Enquist and Arak's (1994) evolutionary search algorithm can be separated into two camps. The first concerns the manner in which they model evolution, the second, the manner in which they model co-evolution.

Evolution

Enquist and Arak choose to model evolution as hill-climbing. A population is represented by one genotype which all members of the population are assumed to share. A mutant genotype is generated and compared to the parent. If assessed as fitter, the mutant genotype is accepted as representing the entire population. This account is motivated by the assumption that mutation events are rare and fitter mutants when they do occur will rapidly dominate a population. This approach to modelling evolution is a natural extension of genetic and evolutionarily stable strategy models within the biological literature in which populations are often considered to be converged upon a particular strategy before the success of individual mutant strategists is assessed. The hill-climbing approach is taken less frequently within artificial life models (although see Kauffman, 1993, for one example) which typically advocate individual based approaches considering populations as aggregates of independent individuals each with their own genotype.

The hill-climbing algorithm employed by Enquist and Arak suffers from problems typical of local search algorithms. Roughly 1 in 20 simulations floundered immediately, as no mutant retina could discriminate the initial signal at above chance, and no mutant signal was discriminated at above chance by the initial retina. Arak and Enquist (1995a, p.340) propose evolutionary-functional explanations for this type of phenomenon rather than explaining it as a consequence of the simplicity of their model evolutive process.

An individual based simulation with a population size of greater than one (i.e., a parallel search algorithm more typical of evolutionary simulation models) reduces the incidence rate of such ‘sterile’ initial conditions, which rapidly falls to near zero as the size of each population increases.

Algorithms with a necessarily converged population, such as the one implemented by Enquist and Arak, also preclude the appreciation of frequency-dependent selection effects, or other effects due to interactions between individuals. Although such effects may not be predicted, their occurrence should not be ruled out *a priori*.

Co-evolution

Enquist and Arak implement co-evolution as a series of consecutive bouts of evolution experienced first by one population, then another. First one population is subjected to a hill-climbing algorithm for a period of evolutionary time, next the second population is similarly evolved, before the process repeats. Since the fitness of each population is crucially dependent upon the other, co-evolution of a sort results.

However, the two-step nature of this co-evolutionary algorithm proves problematic. As was stated in section 6.3, precautions must be taken in order to avoid preferentially evolving either network or signal. In addition, the arbitrary length of each evolutionary ‘step’ (i.e., the value given to *n* in this chapter) is a free parameter, the effect of which is difficult to predict or explore effectively.

Generally, it seems unlikely that one population will remain static throughout the other population's 'generation' of evolutionary time, given that each such 'generation' involves a large number of evolutionary steps, each of which is assumed to be of a length sufficient for a rare mutant to (i) appear, and (ii) if fitter than its parent, fixate in the population. On top of this, as discussed in the passage on noise (section 6.7.3), each fitness assessment is considered to take place over a long (technically infinite) period.

6.7.5 Summary

These methodological issues revolve around subtle characteristics of evolutionary algorithms. These characteristics are still being researched, and the ramifications of differing simulation designs are still being explored. No doubt novel techniques which circumvent some of the problems currently endemic within simulation design will be forthcoming. For the meantime, appreciation of the issues raised above and the growing literature concerning adaptive algorithms and their uses (e.g., Mitchell, 1996) is paramount if modelling work is to proceed successfully.

6.8 Conclusion

In conclusion, it has been shown herein that the form of the signals that Enquist and Arak (1994) evolved was dependent on the presentation regime that the signals underwent. The signals exploited the structure inherent in the presentation regime rather than any 'hidden preferences' inherent in the artificial retinae. Enquist and Arak's unnatural regime was relaxed so as to more closely approximate that experienced by natural signals. As a result, the interesting structure of their artificial patterns dissolved until bland homogeneity was reached. Such homogeneous patterns fail to support the hypothesis that complex symmetry "may arise as a by-product of the need to recognise objects irrespective of their position and orientation in the visual field" (ibid., p.169). Furthermore, a sensory bias for homogeneity such as the one considered here would actively resist the evolution of complex patterns (whether symmetrical or not). Despite this negative conclusion alternative hypotheses linking the form of complex patterns to facts about sensory architecture remain healthy (see Osorio, 1996, for an example of this kind of hypothesis).

Chapter 7

The Evolution of Honest Signals

Hurd's (1995) model of a discrete action-response game, in which the interests of signallers and receivers conflict, is extended to address games in which, as well as signal cost varying with signaller quality, the value of an observer's response to a signal is also dependent on signaller quality. It is shown analytically that non-handicap signalling equilibria exist for such a model.

Using a distributed Genetic Algorithm (GA) to simulate the evolution of the model over time, the model's sensitivity to initial conditions is explored, and an investigation into the attainability of the analytically derived Evolutionarily Stable Strategies (ESSs) is undertaken. It is discovered that the system is capable of attaining signalling equilibria in addition to those derived via analytic techniques, and that these additional equilibria are consistent with the definition of conventional signalling offered in chapter 5.

Grafen's (1990b) proof of Zahavi's handicap principle is generalised in an analogous manner. It is demonstrated analytically that non-handicap signalling equilibria also exist for this continuous model of honest signalling. In addition an evolutionary simulation model is used to corroborate this result, and to explore the effect of varying initial conditions on the attainability of honest signalling equilibria.

7.1 Introduction

In the wake of the fall of group-selectionist thought during the mid-sixties, theoretical biologists were faced with many problems which had previously been comfortably dealt with through some appeal to the worth of behaviours at a group level. The existence of stable signalling systems was one such problem. Although it was feared that the selfish actions of individuals might compromise the stability of natural signalling systems, such systems appeared to be the frequent products of evolution. In the mid-seventies, Zahavi (1975, 1977a) proposed that the stability of such signalling systems may be maintained by a 'handicap principle', i.e., that the differential costs paid by signallers of differing quality might ensure that honest advertisement is an Evolutionarily Stable Strategy (ESS). The reasoning runs something like this…

> "If signallers differ in some variable of interest to an observer (let's call it quality), observers will be selected to take advantage of any honest indicator of this quality. A signal made as an advertisement of quality will necessarily incur some cost. If, for any signal, high quality signallers suffer less production costs than low quality signallers, then signallers are able to demonstrate their true quality through advertising

> more strongly than their poorer competitors. Once this strategy is adopted by the signalling population, the signal is an honest indicator of underlying quality. It cannot be invaded by cheats because to signal more strongly than your quality dictates results in a production cost which is not compensated for by the observer response."

However, a parallel argument runs something like this. . .

> "If signallers differ in some variable of interest to an observer (let's call it need), observers will be selected to take advantage of any honest indicator of this need. A positive response made to an advertisement of need will necessarily induce some benefit. If, for any observer response, high need signallers gain more benefit than low need signallers, then signallers are able to demonstrate their true need through advertising more strongly than their less needy competitors. Once this strategy is adopted by the signalling population, the signal is an honest indicator of underlying need. It cannot be invaded by cheats because to signal more strongly than your need dictates can only result in a response which is not worth enough to compensate the increased production cost."

Notice that whilst the former argument (e.g., Enquist, 1985; Grafen, 1990b; Hurd, 1995) assumes differential costs (i.e., that signaller quality might, to some extent, affect the cost of signal production), the latter does not, and that whilst the latter argument (e.g., Godfray, 1991; Maynard Smith, 1991) assumes differential benefits (i.e., that signaller quality might, to some extent, affect the worth of an observer's response), the former does not.

The former argument might be used to support claims that stotting gazelles are honestly informing predators of their ability to outrun a potential pursuer (e.g., Hasson, 1991; Vega-Redondo & Hasson, 1993). Similarly, the latter argument might be used to support claims that begging nestlings are honestly informing their parents of their need for food items.

Godfray (1991) has provided just such an argument for offspring begging calls. He demonstrates that honest signals of offspring need may be ensured by the facts that (i) signals are costly (he assumes that signal costs are constant across offspring irrespective of their need), and that (ii) the worth of parental resources increases with offspring need (i.e., differential benefits but no differential costs). In Godfray's model, parents are selected to respond positively toward offspring with high need.

Grafen (1990b) considers a similar situation, but with differing assumptions. He suggests that honest signals of offspring quality might be ensured by the facts that (i) parental resources are valuable (he assumes that either resource value is constant across offspring irrespective of their quality, or that resource value increases with offspring quality), and that (ii) the cost of signalling decreases with offspring quality (i.e., differential costs and constrained differential benefits). In Grafen's model, parents are selected to respond positively toward offspring with high quality.

Johnstone (1997) has characterised these two accounts as belonging to two differing classes of model. The first class attempts to account for the evolutionary stability of the honest advertisement of need as a result of the manner in which signaller benefits vary with need. The second class attempts to account for the evolutionary stability of the honest advertisement of quality as a result of the manner in which signaller costs vary with quality.

Here, attention will be drawn to the fact that costs are not absent from scenarios involving the advertisement of need, and similarly, benefits are not absent from scenarios involving the

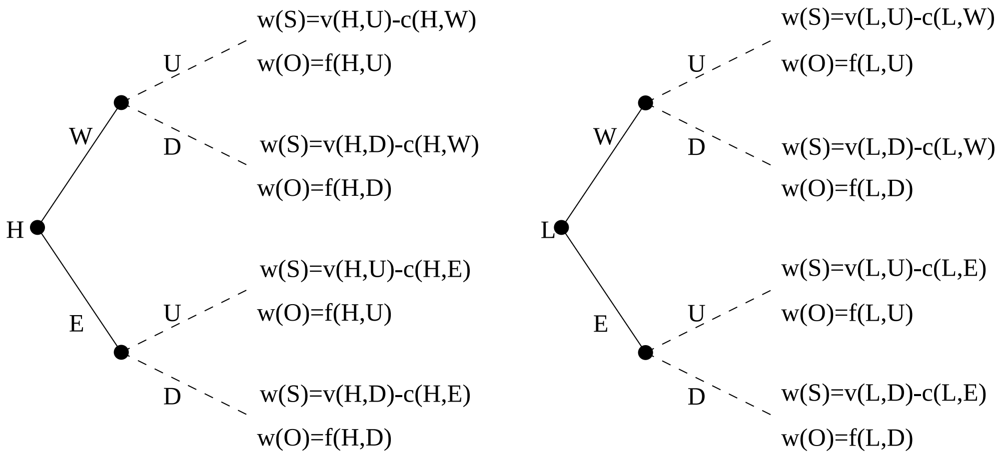


Figure 7.1: Decision trees and fitness consequences for a discrete action-response game. Initially, a Signaller (S) finds itself in one of two possible states (H or L) depicted by the two leftmost nodes of the decision trees. S makes one of two possible signals (E or W) depicted by a labelled solid line. Subsequently an Observer (O), naive as to the state of S, but informed by S's signal, makes one of two possible responses (U or D) depicted by a labelled dashed line. The fitness consequences of each of the eight possible interactions are depicted at the terminal node of each branch of the two decision trees. See text for further clarification.

advertisement of quality. In addition to being present within such scenarios, both factors may vary with either quality or need, and indeed each may vary with both.

In the following sections a simple discrete game, originally due to Hurd (1995), is extended to explore the effects upon signalling equilibria of including, within a signalling model, the impact of both differential costs and differential benefits upon signaller fitness. Section 7.2 will detail the basic game and the simple extension to it. Section 7.3 will describe an implementation of the game as an evolutionary simulation model. Section 7.4 will discuss the relationship between quality and need, and demonstrate that, within the discrete model considered here, only under certain conditions is the honest advertisement of both quality and need possible. Section 7.5 will review and extend Grafen's (1990b) continuous signalling game, whilst section 7.6 will present an implementation of this game as an evolutionary simulation model. Section 7.7 will discuss the relation between the analytic and simulation results presented here, and those presented by Grafen (1990b), Godfray (1991) and Hurd (1995). It will be concluded that meeting Zahavi's two handicap conditions is neither necessary nor sufficient for the existence of an honest communication ESS.

7.2 A Discrete Signalling Game

Hurd (1995) described a game in which a Signaller (S) is privy to some secret (either High or Low) which is of interest to an Observer (O). S makes a signal (East or West) to O. O, in return, makes a response (Up or Down) of interest to S. The game is schematised in figure 7.1.

A signalling strategy determines which signal to make in each of the two states. There are exactly four such strategies. Similarly a response strategy determines which response to give

Signalling Strategies and Response Strategies				
Bit Pattern	Signalling Strategy		Response Strategy	
(0,0)	S(East,East)	Cynic	O(Down,Down)	Mean
(0,1)	S(East,West)	Honest	O(Down,Up)	Believer
(1,0)	S(West,East)	Liar	O(Up,Down)	Non-Believer
(1,1)	S(West,West)	Bluffer	O(Up,Up)	Generous

Table 7.1: Each of the four possible Signalling Strategies, depicted in the form S(what to do if state is Low, what to do if state is High), and four possible Response Strategies, depicted as O(what to do if S plays East, what to do if S plays West), with their associated bit-pattern and descriptive. See section 7.3 for the rationale underlying the allocation of descriptive terms to strategies.

Signalling Strategy-Response Strategy Pairs				
	Response Strategy			
Signalling Strategy	O(Up,Up)	O(Up,Down)	O(Down,Up)	O(Down,Down)
S(East,East)	.	.	.	.
S(East,West)	.	a	b	.
S(West,East)	.	c	d	.
S(West,West)	.	.	.	.

Table 7.2: Each of the four possible Signalling Strategies and Response Strategies are shown. The four Signalling-Strategy/Response-Strategy pairs which constitute communication (*sensu* Enquist, 1985) are denoted *a* through *d* whilst non-communicative pairings are denoted with a period.

to each signal. There are four such response strategies (see table 7.1). Under Enquist's (1985) definition of communication, only four of the 16 possible signal-strategy/response-strategy pairs constitute communication, as only these four prescribe different signals in response to different Signaller states, and different Observer responses to these different signals. This is represented schematically in table 7.2.

The fitness consequences of moves in this discrete action-response game will follow those defined by Hurd (1995). In addition, and in contrast, to Hurd's model, we will assume that the value, to a Signaller, of an Observer's response to a signal is *not* independent of the Signaller's initial state.

Signaller fitness, w_S, is calculated as the cost of signalling subtracted from the benefit derived from the Observer response. The former term is defined as a function, c, of the Signaller's initial state, I (either High or Low), and the signalling action, A (either East or West), whilst the latter is defined as a function, v, of the Signaller's initial state, and the Observer's response, R (either Up or Down),

$$w_S = v(I,R) - c(I,A).$$

Similarly, Observer fitness, w_O, is calculated as a function, f, of the state of the Signaller, and the Observer response,

$$w_O = f(I,R).$$

The fitness consequences of each of the eight possible signalling scenarios are depicted in figure 7.1.

Hurd defines the payoffs in order that the interests of S and O conflicted. Observers benefit from responding Up to High-state Signallers, and Down to Low-state Signallers,

$$w_O(H,U) > w_O(H,D),$$

$$w_O(L,U) < w_O(L,D),$$

whilst Signallers benefit from eliciting an Up, rather than a Down, response from Observers,

$$v(H,U) > v(H,D),$$

$$v(L,U) > v(L,D).$$

After Hurd, we define the relative value of an Up response for each class of Signaller as

$$V_H = v(H,U) - v(H,D) > 0,$$

$$V_L = v(L,U) - v(L,D) > 0.$$

Similarly, we define the relative cost of signalling West for each class of Signaller as

$$C_H = c(H,W) - c(H,E),$$

$$C_L = c(L,W) - c(L,E).$$

In order that S(E,W) be the unique, best response to O(D,U) (the Signalling-Strategy/Response-Strategy pair arbitrarily chosen by Hurd as a candidate ESS), it must be the case that,

$$v(H,U) - c(H,W) > v(H,D) - c(H,E),$$

$$v(L,D) - c(L,E) > v(L,U) - c(L,W).$$

By substitution, it follows that,

$$V_H > C_H,$$

$$V_L < C_L.$$

It is plain that Hurd's result, $C_L > V > C_H$, is the special case inequality resulting from the substitution of $V = V_H = V_L$, i.e., the assumption that "V is equal for all signallers"(Hurd, 1995, p.219). Hurd depicts his special case graphically (see figure 7.2a). He points out that signalling equilibria exist in part of the region of the graph defined by $C_H \leq 0$, which he interprets as indicating that 'handicap' signals need not be costly for High-state Signallers at equilibria, and indeed may be given preferentially by High-state Signallers. He also points out that despite the fact that all signalling equilibria satisfy the inequality $C_L > C_H$, signalling equilibria do not exist in certain

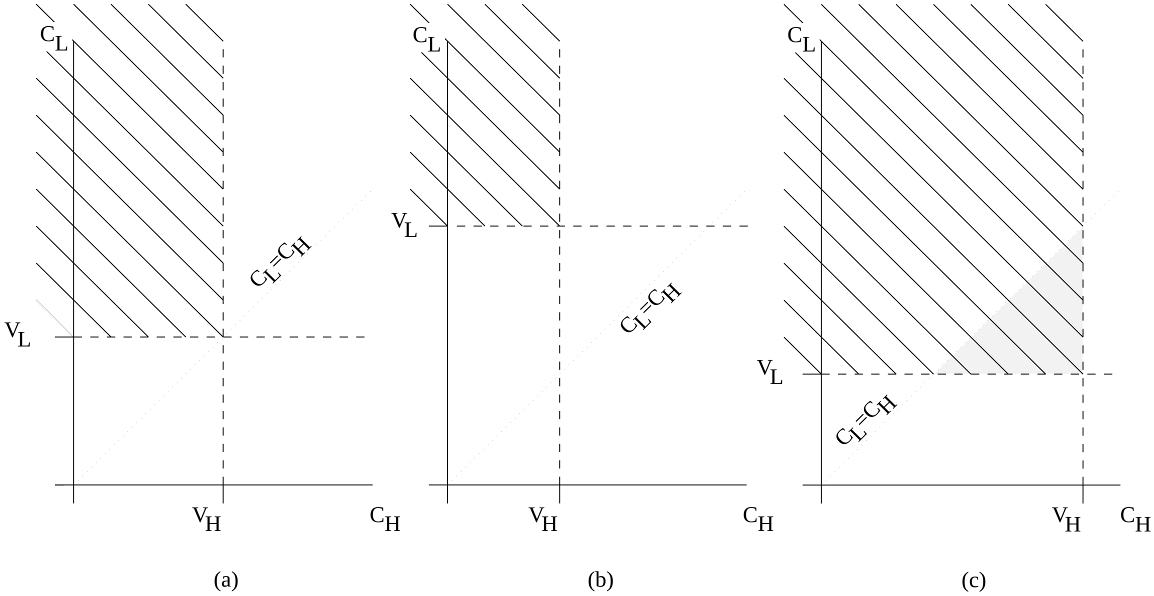


Figure 7.2: In each graph a pair of cost parameters (C_H, C_L) specifies a point in the plane of all possible versions of the discrete action-response game for a particular pair of value parameters (V_H, V_L) which divide the space into four quadrants. Graphs depict *(a)* Hurd's (1995) result in which $V_H = V_L$, *(b)* a scenario in which $V_H < V_L$, and *(c)* a scenario in which $V_H > V_L$. In each graph the diagonal hatching corresponds to (C_H, C_L) parameter values which afford stable communication equilibria, the line $C_L = C_H$ divides the space into two areas, the upper of which is predicted, according to Zahavi's argument, to contain handicap equilibria, whilst the lower is predicted to offer no communication equilibria. The shaded area in *(c)* highlights non-handicap parameter values in which (*contra* Zahavi) stable signalling may occur.

areas of the graph satisfying this inequality, i.e., that $C_L > C_H$ is necessary but not sufficient for communication to be stable.

However, under conditions modelled here, in which $V_L \neq V_H$, it can be shown that Zahavi's handicap principle is neither necessary *nor* sufficient for the existence of signalling equilibria (see figure 7.2b and c). When the value of a beneficial response is greater for Low-state Signallers than High-state Signallers (i.e., $V_L > V_H$, see figure 7.2b) signalling equilibria lie above the line $C_L = C_H$, but when the value of a beneficial response is higher for High-state Signallers (i.e., $V_L < V_H$, see figure 7.2c) signalling equilibria may lie below the line defined by this inequality.

7.3 An Evolutionary Simulation Model of a Discrete Signalling Game

Whilst the analysis presented in the previous section reveals which areas of the parameter space admit of honest evolutionarily stable strategies (ESSs), it makes no claims concerning admissible trajectories in the state space occupied by a population of signallers and receivers playing a particular version of this discrete action-response game. In addition, the analysis above makes no attempt to describe the behaviour of systems which fail to attain an honest signalling ESS. Little attention has been paid to how honest signalling might evolve from non-signalling behaviour, although Yachi (1995) has attempted to characterise the conditions under which handicap signalling might evolve.

Therefore, in order to discover empirically whether signalling equilibria are attainable by a population initially behaving ‘randomly’, and to explore the behaviour of the system prior to (potentially) achieving an honest signalling ESS, an evolutionary simulation modelling approach was undertaken[1].

Two populations (one of signallers, one of receivers) were distributed across the same 25-by-25 grid world. Each cell in the grid contained one signaller and one receiver. Each signaller was allocated a discrete internal state (either High or Low) at random, i.e., the internal trait was non-heritable[2]. In addition, each signaller inherited one of the four possible signalling strategies (represented as a two-bit binary number) from its parent. Similarly each receiver inherited one of the four possible response strategies (again represented as a two-bit binary number) from its parent (see table 7.1). The fitnesses of signallers and receivers were calculated as shown in figure 7.1, each signaller interacting once with the receiver sharing its cell.

Once each signaller and receiver had been assessed the whole population was updated synchronously. The location of a potential parent was chosen through perturbing both the x and y grid co-ordinates of the offspring cell by independent values drawn from a normal probability distribution with standard deviation 0.75 and mean zero. Six potential parents were chosen for each offspring signaller. An offspring signaller inherited its signalling strategy from the fittest of these six. Similarly, an offspring receiver inherited its response strategy from the fittest of six receivers independently chosen from the previous generation in the same manner. A mutation rate of one bit in one hundred ensured that offspring sometimes inherited a strategy which differed from that of their parents. Populations were simulated for 500 generations in this manner, during which time the proportions of signallers playing each of the four possible signalling strategies, and the proportions of receivers playing each of the four possible response strategies, were recorded.

In order to fully specify a simulation run, several parameter values must be decided upon. The costs of signalling each of the two possible signals (East or West) must be specified for each of the two possible signaller states (High or Low). Similarly, the benefit of obtaining each of the two possible responses (Up or Down) must be specified for each of the two possible signaller states. Finally the value to the receiver of making each of the two possible responses must be specified for each of the two possible signaller states.

The fitness consequences of receiver responses *for the receiver* were fixed at 40 for responding Up to a High-state signaller, or Down to a Low-state signaller, and zero otherwise.

The cost of signalling East for both Low-state signallers and High-state signallers was fixed at zero. All 576 possible pairs drawn from the set $\{10.0, 12.5, 15.0, ..., 70.0\}$ were explored as costs of signalling West for High-state signallers, and signalling West for Low-state signallers.

---

[1] Copies of the code, and a version of this chapter with colour figures, are available from the world-wide-web page `http://www.cogs.susx.ac.uk/users/sethb/ecal97.html`

[2] This is in accordance with many models of the evolution of signalling (e.g., Hurd, 1995; Grafen, 1990b). Models in which the advertised trait, in addition to the advertising strategy, is itself heritable encounter a problem known within evolutionary theory as the lek paradox. A full account of this problem is beyond the scope of this chapter (interested readers are directed to Kirkpatrick & Ryan, 1991; Pomiankowski & Møller, 1995). Briefly, in simple models of signal evolution involving a heritable advertised trait, the variability of the trait across the population tends to decrease over evolutionary time. As the variation in the trait falls observers find any signal which distinguishes between signallers with differing traits less and less informative. As a consequence signalling (which involves some cost to the signaller and, possibly, the observer) tends to die out.

The value *to a signaller* of a Down response was fixed at zero for both High- and Low-state signallers. The value *to a signaller* of a receiver response Up was drawn from the set {(40,40), (50,30), (30,50)} where the figures in parentheses denote (value to Low-state signaller, value to High-state signaller). These three pairs can be represented by figures 7.2a, 7.2b, and 7.2c respectively.

These parameter values allow the exploration of cost parameters lying in each of the four quadrants for each of the three classes of scenario depicted in figure 7.2.

The rationale underlying the choice of labels used throughout the results section to describe the possible strategies (see table 7.1) reflects the costs and benefits described above. Signalling East is a costless action and is thus the default signalling behaviour, whereas signalling West is costly and will be regarded as a positive action in comparison. Thus a signaller which always signals West will be dubbed a ‘Bluffer’, whilst one which signals West only when High state will be described as ‘Honest’ in that a positive signal is being used to advertise a positive (High) trait. Similarly, as obtaining a Down response is not beneficial to signallers, receivers which always respond Down will be termed ‘Mean’ in comparison to ‘Generous’ strategists which always respond Up.

The initial conditions imposed upon the populations were also varied. Populations initially with random behaviour (strategies drawn at random from the strategy set), were compared to populations initially converged at an Honest signalling strategy and Believing response strategy, and populations initially converged at a Cynical signalling strategy and Mean response strategy. These three classes of initial conditions will hence forward be referred to as ‘Random’, ‘Honest’, and ‘Cynical’ initial conditions, respectively.

7.3.1 Results

For each setting of the value parameters, a pair of cost parameters was taken to specify a system lying within one of four quadrants defined by the two inequalities

$$V_H > C_H,$$

$$V_L < C_L.$$

From the analysis carried out in section 7.2, systems residing in the top-left quadrant of parameter space (hereafter Quadrant 1) satisfy the conditions for the existence of an honest signalling ESS. Systems residing in the top-right quadrant (hereafter Quadrant 2) cannot support honest communication as the costs of signalling are too great for both High- and Low-state signallers. Systems residing in the bottom-left quadrant (hereafter Quadrant 3) cannot support honest communication as the costs of signalling are bearable for signallers of Low state allowing them to mimic High-state signallers. Systems residing in the bottom-right quadrant (hereafter Quadrant 4) cannot support honest signalling as High-state signallers cannot afford to signal, whilst Low-state signallers can.

Five classes of behaviour were exhibited by the system. Stereotypical examples of trajectories through strategy space for four of these classes are presented in figure 7.3, whilst the remaining behaviour is plotted in figure 7.4[3]. The distribution of these five classes of behaviour across parameter space is represented by figure 7.5.

---

[3] This trajectory is plotted separately to avoid confusing figure 7.3, and to highlight the variation in population makeup it involves.

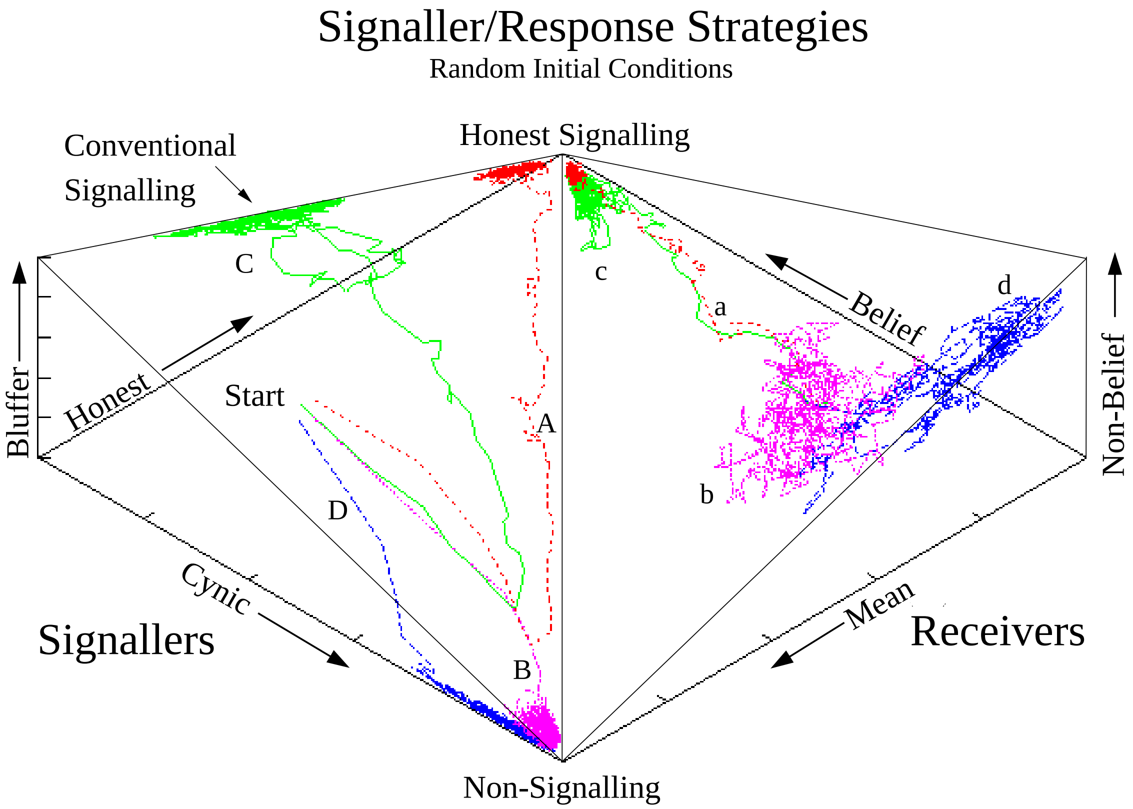


Figure 7.3: Stereotypical trajectories through strategy-space for four of the five classes of behaviour exhibited by the model. Populations were evolved from Random initial conditions. The left half of the plot contains trajectories representing the change in strategic make-up experienced by particular populations of signallers, whilst the right half contains similar trajectories for populations of receivers. The state of a population at any point can be represented by a four dimensional vector, each element of which denotes the proportion of the population adopting one of the four strategies. Such vectors, being of unit sum, have only three degrees of freedom, and may be naturally plotted as points in a 3-d wedge-shaped space. Distance from the origin along each of the axes indicates the proportion of the population adopting the strategy denoted by the axis label, with the remaining fourth strategy implicit in the graph (decreasing with distance from origin). Associated pairs of signaller and receiver trajectories are denoted by the same letter (upper case denotes signaller trajectories, lower case denotes receiver trajectories).

Trajectory (*a*): *Honesty* is produced only by systems with Quadrant 1 parameters; populations converge on Honesty and Belief. This class of behaviour corresponds to the honest signalling ESS predicted in section 7.2. Although this ESS existed for all games within Quadrant 1 (i.e., from Honest initial conditions, no simulation ever deviated from Honesty), simulations from Random initial conditions, with parameters for which the inequality $V_H > V_L$ held, often failed to reach it.

Trajectory (*b*): *Non-Signalling* is found in Quadrants 2 and 4 for parameter values satisfying the inequality $V_H \leq V_L$; signalling populations converge on Cynic, whilst receiver populations wander in the centre of strategy space. Within this class of behaviour any strategy adopted by the receiver population can be exploited by the Low-state signallers, thus no clear response strategy

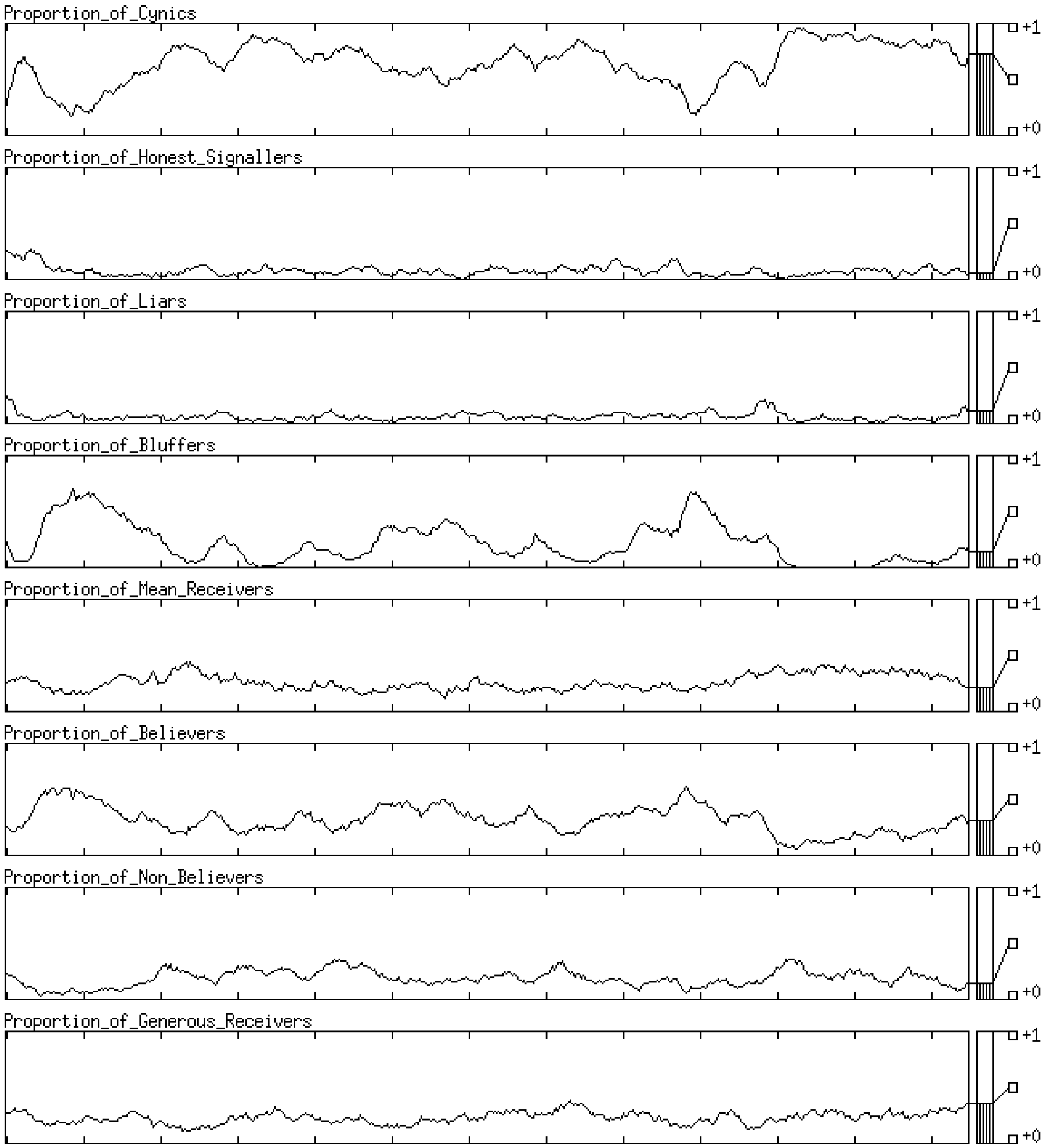


Figure 7.4: A stereotypical *conventional non-signalling* trajectory through strategy-space. Populations were evolved from Random initial conditions. The upper four plots contain trajectories representing the change in strategic make-up experienced by a particular population of signallers, whilst the lower four contain similar trajectories for the associated population of receivers. Each of the four plots represents the changing proportion of each population adopting one of the four strategies available.

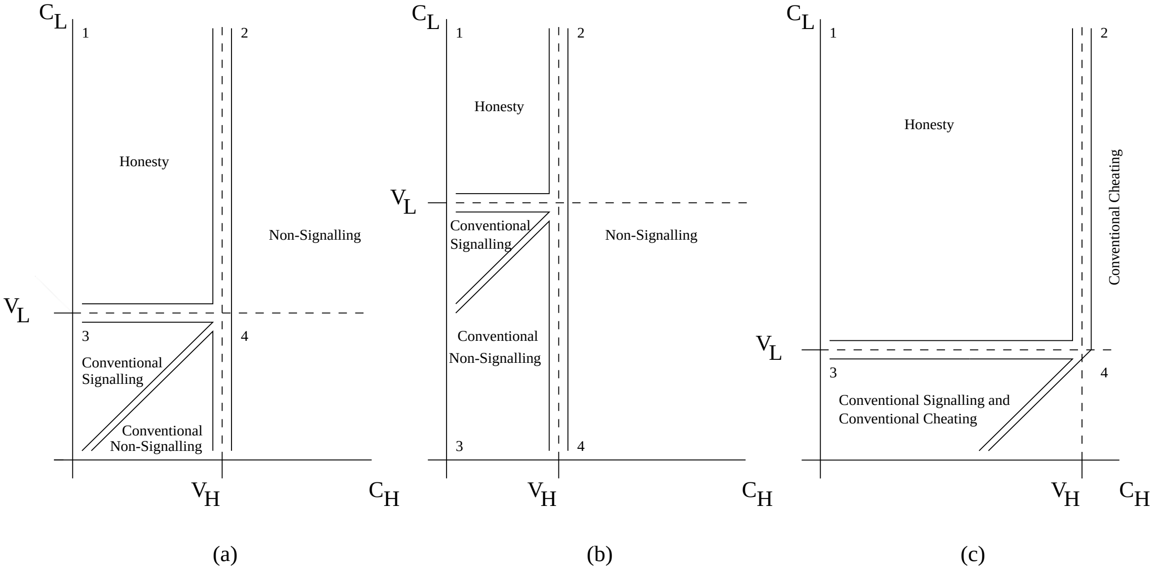


Figure 7.5: Graphs as per figure 7.2 showing the classes of behaviour observed across the parameter space of the extended discrete action-response model. Honest behaviour is confined to the predicted quadrant of parameter space for all three graphs. See text for a full discussion.

emerges, and signallers cut their losses by refusing to signal.

Trajectory (c): *Conventional Signalling* is found only for parameter values lying within Quadrant 3, and satisfying the inequality $V_H - C_H > V_L - C_L$; signalling populations converge on a fluctuating mixture of Honesty and Bluffing, whilst receiver populations converge on Belief but maintain a significant, but very low (and fluctuating) frequency of both Mean and Generous strategists. This class of behaviour is a conventional signalling scenario suffering a degree of Bluffing strategists who exploit Believing receivers. Bluffing cannot come to dominate the signalling population however, since as the rate of Bluffing increases, the fitness of the Believing strategy, upon which the fitness of Bluffers depends, decreases. This negative feedback between the signalling and receiving populations results in oscillations within each population.

The stability of the remaining two classes of behaviour is also maintained through weak negative feedback interactions which induce cyclic trajectories typical of conventional signalling scenarios.

Trajectory (d): *Conventional Cheating* is found only for games in which $V_H > V_L$. For such games, this class of behaviour can be stable for any pair of cost parameters; signalling populations converge on Cynic with a fluctuating proportion of Liars, whilst receiver populations are mixed, but feature an over-representation of Non-Believers. This class of behaviour is a non-signalling scenario suffering a low level of Liars. As the frequency of lying increases, the proportion of Non-Believers increases in the receiver population, thus reducing the fitness of the Liar strategy which relied on Generous or Believing positive responses to the Low-state signallers, in order to out-perform Cynics, which, as a result, increase in numbers.

The fifth class of behaviour: *Conventional Non-Signalling* is found only within Quadrant 3, under parameter values satisfying the inequalities $V_H - C_H < V_L - C_L$, and $V_H \leq V_L$; signalling populations converge on Cynic with regular insurgences of Bluffing strategists whilst receiver populations wander in the centre of strategy space with a slight over-representation of Believers. The invading Bluff strategy exploits the over-representation of Believing receivers, but is prevented from dominating the signaller population by negative feedback from the receiver population.

To summarise, several interesting, robust phenomena, which were opaque to the analysis carried out in section 7.2 have been detailed. The behaviour of this very simple system varies from non-signalling equilibria, through scenarios in which stability is maintained through reciprocal fitness interactions which constitute the negative feedback indicative of conventional signalling (see chapter 5 herein), to honest signalling equilibria in which honesty is maintained though the interaction of differential costs and benefits. Further exploration of the system's behaviour will be necessary before the factors governing the mode of behaviour that will evolve in a particular case can be made more explicit.

However, the discrete nature of the action-response game considered here, although attractive through its tractability, also risks lacking application to natural signalling through this very simplicity. Do the classes of behaviour exhibited in a discrete game such as the one considered above exist for more complex models? As a first step towards answering this question, an analysis of Grafen's (1990b) model is undertaken in an effort to demonstrate that at least the results derived analytically in section 7.2 will generalise to a continuous model.

However, before constructing such a model, some discussion of the relationship between quality and need within signalling models is necessary.

7.4 Quality and Need

As was mentioned in the opening passage of this chapter, two complementary accounts of the handicap principle have been constructed. One of these accounts addresses the advertisement of quality, in which honesty is stabilised by differential signalling costs; the other addresses the advertisement of need, in which honesty is stabilised by differential benefits. Since the model constructed above involves both differential costs and differential benefits, it offers the possibility of modelling both the advertisement of need and the advertisement of quality, and the potential to explore the relationship between these two concepts.

7.4.1 Quality and Need Unrelated

Until now, signaller state has been referred to as either High or Low. Whether these labels refer to high and low quality, or high and low need, has been left unspecified. As it stands, the model is neutral with respect to this distinction. One may freely interpret signaller state as indicating either need or quality. The conditions for the existence of an honest signalling ESS remain unchanged.

Under one reading, signallers in a High state are high quality signallers, whilst signallers in a Low state are low quality signallers. Under such a reading, the conditions for the existence of an honest signalling ESS, in which, arbitrarily, high quality signallers exclusively signal West, whilst low quality signallers exclusively signal East, can be represented by two inequalities, thus,

$$V_H > C_H,$$

$$V_L < C_L.$$

Under the second reading, signallers in a High state are high need signallers, whilst signallers in a Low state are low need signallers. Under such a reading, the conditions for the existence of an honest signalling ESS, in which, arbitrarily, high need signallers exclusively signal West, whilst low need signallers exclusively signal East, can be represented by the same two inequalities.

Thus the model is clearly one which addresses the advertisement of an arbitrary property. In addition, the results of the model suggest that to claim that honest advertisements of quality must be maintained through differential costs (i.e., $C_H > C_L$), or that honest advertisements of need must be maintained through differential benefits (i.e., $V_H > V_L$), is to offer an incomplete account. Honest advertisements of quality may be maintained through differential benefits (given certain constraints on signal costs), and similarly, honest advertisements of need may be maintained through differential costs (given certain constraints on benefits). The nature of this inter-dependency of constraints is captured by the two inequalities which govern the honest advertisement of any binary trait under the conditions modelled by the discrete action-response game outlined above.

7.4.2 Quality and Need Related

However, the account presented above is predicated upon notions of need and quality as independent traits, i.e., modelling need was treated as separate from modelling quality. Such models cannot capture the notion of relatedness between need and quality which appears to drive the construction of the two complementary accounts of the handicap principle with which this chapter opened.

Typically, since each model of advertising deals either with quality *or* need, little space is devoted to the relationship between these concepts. However, although the relationship between need and quality is not explicit in any of the models which deal with the advertisement of either trait, we may postulate some interaction between the two concepts which seems to accord with their general usage.

For example, since quality and need are typically dealt with as symmetrically opposed concepts, we might propose that quality be modelled as varying inversely with need. This notion seems to underly the construction of Godfray's (1991) model of offspring begging, in which the lower a chicks quality (i.e., the lower the chick's expected number of offspring, or chance of reaching maturity, etc.) the greater its need for resources, and *vice versa*.

The construction of a single continuum along which need varies inversely with quality has interesting implications for signalling models, which help to highlight the role played by the observer. For example, under this reading of the relationship between quality and need, any advertisement of quality will simultaneously be a signal of need. One might therefore imagine that in situations in which, for example, the honest advertisement of quality was an ESS, the honest advertisement of need would also result. However, as we shall discover, this is sometimes not the case.

ESSs for Need and Quality

If need is assumed to vary inversely with quality in some manner, two classes of model may be characterised. Models of the advertisement of need may be distinguished from models of the advertisement of quality on the basis of the interests of the observing party.

First, an interpretation of signaller state must be settled upon in some arbitrary manner. For this analysis we will stipulate that High-state signallers are those of high quality and low need, whilst Low-state signallers are those of low quality and high need.

Secondly, an interpretation of the worth of observer responses must be fixed upon. Following the convention observed throughout the analysis above, the Up response will be deemed more valuable than the Down response, irrespective of signaller state (although, of course, the magnitude of the difference between the values of the two responses may itself differ with signaller state),

$$V_H = v(H,U) - v(H,D) > 0,$$

$$V_L = v(L,U) - v(L,D) > 0.$$

The relative cost of signalling West for each class of Signaller also remains unchanged,

$$C_H = c(H,W) - c(H,E),$$

$$C_L = c(L,W) - c(L,E).$$

With this interpretation in place, the interests of the receiver may be manipulated in order to distinguish between models of the advertisement of need, and models of the advertisement of quality.

For models of the advertisement of quality, the conditions governing observer fitness are defined by

$$w_O(H,U) > w_O(H,D), w_O(L,U) < w_O(L,D). \tag{7.1}$$

These inequalities specify that observers are selected to respond Up to High-state (i.e., high quality) signallers, and Down to Low-state (i.e., low quality) signallers.

Conversely, for models of the advertisement of need, the conditions governing observer fitness are defined by

$$w_O(H,U) < w_O(H,D), w_O(L,U) > w_O(L,D). \tag{7.2}$$

These inequalities specify that observers are selected to respond Up to Low-state (i.e., high need) signallers, and Down to High-state (i.e., low need) signallers.

As depicted in table 7.1, for the discrete action-response game being considered here, there are four possible pairs of signaller strategy and response strategy which are candidates for communicative signalling ESSs, denoted *a* through *d*.[4]

Two of these possible ESSs (*b* and *c*) only exist for models of the advertisement of quality since they correspond to scenarios in which observers respond positively (Up) to High-state (i.e., high quality) signallers, i.e., they are stable under inequalities (7.1).

The remaining pair (*a* and *d*) only exist for models of the advertisement of need since they correspond to scenarios in which observers respond positively (Up) to Low-state (i.e., high need) signallers, i.e., they are stable under inequalities (7.2).

---

[4] N.b., these candidate ESSs should not be confused with the five classes of trajectory discussed in the previous section.

For each of these pairs of candidate ESSs there exists one in which signalling West elicits the preferred (Up) response (b for advertising quality, d for advertising need), and one in which signalling East elicits the preferred response (c for advertising quality, a for advertising need).

We are now in a position to ask two questions. First, within each signalling scenario (i.e., the signalling of need, or the signalling of quality), are there conditions under which both (i) signaller state may be honestly advertised, and (ii) the relationship between signal and signaller state is arbitrary? That is, are there conditions under which, for example, using West to signal High state and East to signal Low state is stable, *and* using West to signal Low state and East to signal High state is *simultaneously* stable. This question can be answered by considering the possibility of a non-empty intersection between the ESS conditions for strategy pairs b and c, or between the ESS conditions for strategy pairs a and d.

In order to explore this possibility, the ESS conditions for each strategy pair must be derived.

In section 7.2 above it was shown that the conditions under which strategy pair b was an ESS could be represented by the inequalities,

$$C_H < V_H,$$

$$C_L > V_L.$$

The ESS conditions for strategy pair c can be calculated in the same manner. It has already been noted that, like strategy pair b, strategy pair c is stable under inequalities (7.1) and is thus a candidate for an ESS in which quality is honestly advertised. In addition, for S(W,E) and O(U,D) to be such an ESS,

$$v(H,U) - c(H,E) > v(H,D) - c(H,W),$$

$$v(L,D) - c(L,W) > v(L,U) - c(L,E).$$

By substitution, it follows that,

$$C_H > -V_H,$$

$$C_L < -V_L.$$

Similarly, both strategy pair a, S(E,W) and O(U,D), and strategy pair d, S(W,E) and O(D,U) are stable under inequalities (7.2) and are thus candidates for ESSs in which need is honestly advertised. In addition, for S(E,W) and O(U,D) to be an ESS,

$$v(H,D) - c(H,W) > v(H,U) - c(H,E),$$

$$v(L,U) - c(L,E) > v(L,D) - c(L,W).$$

By substitution, it follows that,

$$C_H < -V_H,$$

$$C_L > -V_L.$$

Finally, for strategy pair d, S(W,E) and O(D,U), to be an ESS,

$$v(H,D) - c(H,E) > v(H,U) - c(H,W),$$

$$v(L,U) - c(L,W) > v(L,D) - c(L,E).$$

By substitution, it follows that,

$$C_H > V_H,$$

$$C_L < V_L.$$

These four sets of conditions, under which each of the strategy pairs *a* through *d* is an honest signalling ESS, are represented graphically in figure 7.6.

From the inequalities governing these four sets of ESS conditions it is plain that $b \cap c = \{\emptyset\}$, and $a \cap d = \{\emptyset\}$, i.e., there exist no conditions on signal cost and response value under which signallers may use either signal to honestly indicate their state.

The second question we are in a position to ask concerns the possibility that scenarios exist in which the same signal might be honestly used to either indicate need (if observers reward need) *or* quality (if observers reward quality). That is, might there be conditions under which a parameterisation of signaller costs and benefits which admits of an honest signalling strategy which advertises quality, also admits of an honest signalling strategy which advertises need? In order to understand this question, we must distinguish between its two possible meanings.

First, this question could be interpreted as ‘What conditions on signal cost and response value must be met in order that both (i) signalling West may be an honest indicator of high quality for scenarios in which receivers reward high quality, and (ii) signalling West may be an honest indicator of high need for scenarios in which receivers reward high need’.

This interpretation of the question concerns the status of the intersection between the ESS conditions for strategy pairs *b* and *d* (or *mutatis mutandis* between those for *a* and *c*).

Secondly this question could be interpreted as ‘What conditions on signal cost and response value must be met in order that both (i) signalling West may be an honest indicator of high quality for scenarios in which receivers reward high quality, and (ii) signalling East may be an honest indicator of high need for scenarios in which receivers reward high need’.

This interpretation of the question concerns the status of the intersection between the ESS conditions for strategy pairs *b* and *a* (or *mutatis mutandis* between those for *c* and *d*).

The first interpretation demands conditions under which a particular signal could be used either to signal quality or need. Using the ESS conditions derived above, it is easy to demonstrate that no such conditions exist.

$$b \cap d = \{\emptyset\}$$

$$c \cap a = \{\emptyset\}$$

There are thus no conditions under which signalling West (or East) could be used both to honestly advertise high quality when receivers reward high quality, and high need when signallers reward high need. From figure 7.6 it can be seen that reversing the interests of the receivers in the game (i.e., changing between inequalities (7.1) and (7.2)) whilst demanding that signalling

West continues to be rewarded is equivalent to reflecting the graph depicted in figure 7.6 about axes defined by $C_H = V_H$, and $C_L = V_L$ (equivalent to a single reflection about the axis $V_L - C_L = V_H - C_H$). Thus, for a game defined by particular choices for C_H, C_L, V_H, and V_L, a change from a scenario in which High-state signallers are favoured by receivers, to one in which Low-state signallers are favoured, results in the movement of the signalling equilibria from Quadrant 1 to Quadrant 3.

The second interpretation of the question raised above demands conditions under which a particular response (e.g., Up) can be honestly be elicited by either High, or Low-state signallers dependent on whether receivers are selected to reward High, or Low-state signallers. A similar analysis to that presented above demonstrates that there are such conditions.

Since, by definition, $V_H > 0$, and $V_L > 0$, the intersection between ESS conditions for strategy pairs *a* and *b* can be represented by the inequalities,

$$C_H < -V_H,$$

$$C_L > V_L,$$

whilst the intersection between ESS conditions for strategy pairs *c* and *d* can be represented by the inequalities

$$C_H > V_H,$$

$$C_L < -V_L.$$

These intersections, depicted in figure 7.6, define scenarios in which both classes of signaller prefer to make different signals. Additionally, each class of signaller prefers one signal to the extent that this preference more than compensates for any incentive to make a different signal in order to obtain a positive observer response. Under such conditions, the agenda of the observer (i.e., which state is rewarded by the observer) is free to change without affecting the signals given by the signallers.

This analysis demonstrates that, for the discrete action-response game considered here, although need and quality are intimately connected (technically they are the same variable), parameter values which admit of honest signalling ESSs in which, for example, West honestly advertises quality, are never parameter values for which (i) East can also be an honest indicator of quality, or (ii) West can be an honest indicator of need, and are only sometimes parameter values in which (iii) East can be used to honestly advertise need.

These relationships between models of the advertisement of quality and the advertisement of need can be further explored in the context of a continuous signalling model due to Grafen (1990b).

7.5 A Continuous Signalling Game

Grafen's (1990b) model of Zahavi's handicap principle upheld Zahavi's contentions that in order for communication to be stable certain relationships between signal cost and signaller quality had to hold. Specifically, the criteria which Zahavi (1975, 1977a) specifies are that (i) signals must

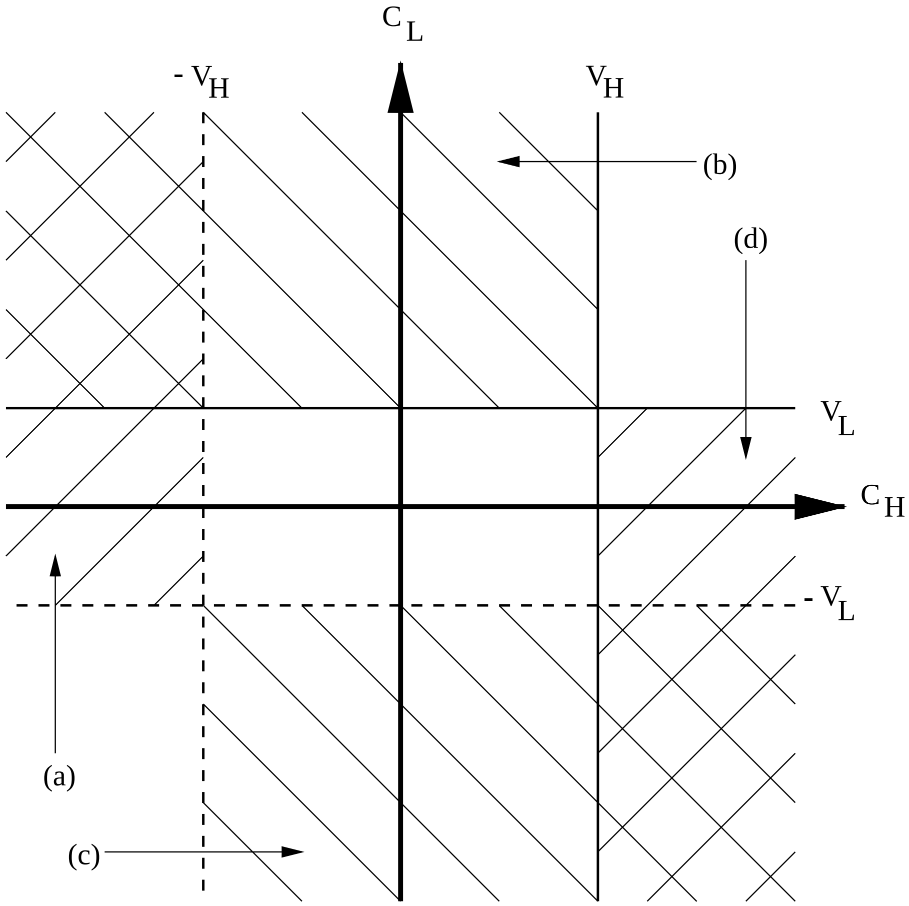


Figure 7.6: Graph as per figure 7.2 showing the areas of the parameter space which admit of an honest signalling ESS for each of the four possible signaller-strategy/response-strategy pairings depicted as per table 7.1. Note that the only non-empty intersections are $a \cap b$, and $c \cap d$. Furthermore, these intersections are defined by conditions under which signallers in one state prefer to make a signal different to that preferred by signallers in the alternative state, (i.e., C_H is of a different sign to C_L) and that this preference is strong enough to prevent a change of signal even if it were accompanied by a positive observer response (i.e., $|C_H| > V_H$, and $|C_L| > V_L$). See text for discussion.

be costly, and that (ii) for any given level of advertisement, signallers of low quality must suffer higher production costs than signallers of higher quality.

Grafen glosses his model in terms of a sexual selection scenario in which males possess, to a certain degree, some non-heritable, continuously varying trait, dubbed quality. The degree to which males possess this trait is of interest to females, who wish to mate selectively with high quality males, but the extent of a male's quality is not directly available to them. Thus, females are selected to respond appropriately to males (i.e., they are selected to respond more positively towards males who possess higher quality), whilst males are selected to elicit maximum positive response. This ensures that signaller and receivers in this continuous game experience a conflict of interest equivalent to that engineered by Hurd (1995).

Under such a gloss 'quality' should be read as indicating not any arbitrary male characteristic, but specifically some male trait which is predictive of male viability, virility, etc. In order to compare Grafen's model with that of Godfray (1991), quality will sometimes be considered to be inversely related to need in an analogous manner to that described within section 7.4.2 above. This will enable a more general continuous signalling model to be considered, in which observers may be rewarded for responding positively to either the possession of quality, *or* its absence, i.e., a model in which the advertisement of both quality, and need, may be addressed. For the moment, however, analysis will concentrate solely on Grafen's model of the honest advertisement of signaller quality.

7.5.1 Grafen's Proof

After defining signaller (male) fitness (w) as a function of three variables, the signaller's level of advertisement (a), the strength of observer (female) preference for advertising (p), and signaller quality (q), Grafen asserted that Zahavi's criteria could be formalised as conditions placed on various partial derivatives of the fitness function. First order derivatives were represented as w subscripted with a digit denoting the variable (a, p, or q) with respect to which the rate of change of fitness was being derived. Second order derivatives were similarly denoted by w subscripted with a pair of digits.

For example, the condition that signals must be costly (i.e., that, as advertising levels increase, fitness decreases) is maintained by the inequality, $w_1 < 0$,

$$\frac{\partial w}{\partial a} < 0.$$

That female preference is beneficial is similarly maintained by the inequality, $w_2 > 0$,

$$\frac{\partial w}{\partial p} > 0.$$

A further condition ensured that "better males do better by advertising more"(Grafen, 1990b, p.520),

$$\frac{\partial w/\partial a}{\partial w/\partial p} \text{ is strictly increasing in q.} \tag{7.3}$$

Grafen demonstrated that if the beneficial fitness consequences of the strength of female preference was independent of signaller quality ($w_{23} = 0$, which Hurd's (1995) model also assumes), or if the beneficial fitness consequences of the strength of female preference were greater for signallers of higher quality ($w_{23} > 0$), then that equation (7.3) holds can be ensured by the maintenance of the following inequality: $w_{13} > 0$ (i.e., that higher quality signallers pay lower advertising costs — Zahavi's second handicap criterion). Grafen proceeds to show that communication equilibria exist under these conditions.

Grafen then attempts to reverse this proof in order to show that *any* stable communication equilibria require that Zahavi's criteria hold, and thus that handicap equilibria are not merely "quirky possibilities"(Grafen, 1990b, p.521).

7.5.2 General Solution

Here, a more general solution to the scheme presented by Grafen will be attempted. Rather than assuming that signaller quality mediates the influence of observer response on signaller quality in certain ways, and then, with these assumptions in place, proceeding to analyse the conditions under which honest advertising of quality is possible, we will explore the full range of possibilities admitted by the model as presented so far. This procedure is analogous to that carried out for the discrete model presented in sections 7.2 and 7.3, and results in analogous findings.

Condition (7.3) can be presented as

$$\frac{\partial(\frac{w_1}{w_2})}{\partial q} > 0$$

which, after application of the quotient rule, can be re-written as

$$\frac{w_{13}w_2 - w_1 w_{23}}{(w_2)^2} > 0.$$

The denominator is necessarily positive, and by assumption, w_1 is negative, whilst w_2 is positive. Thus, discarding the denominator, and dividing through by w_2 casts the general solution to equation (7.3) as

$$w_{13} + w_{23} * |\frac{w_1}{w_2}| > 0. \tag{7.4}$$

We will now explore the form that this inequality takes under each of the three classes of condition governing the manner in which the beneficial effects of signalling for the signaller are moderated by the signaller's quality; an analysis analogous to that carried out in section 7.2 for the extension of Hurd's (1995) model.

First, under the condition in which the beneficial fitness consequence of female preference is independent of signaller quality (i.e., $w_{23} = 0$, analogous to Hurd's $V = V_H = V_L$), equation (7.4) reduces to $w_{13} > 0$. This is Grafen's result (i.e., Zahavi's second handicap criterion).

Under the condition in which the beneficial fitness consequences of female preference are higher for poorer quality signallers (i.e., $w_{23} < 0$), equation (7.4) reduces to,

$$w_{13} > |w_{23}| * |\frac{w_1}{w_2}|.$$

It is plain that, whilst this inequality *requires* that $w_{13} > 0$, it remains the case that the satisfaction of $w_{13} > 0$ is not *sufficient* for signalling to be stable. Lower quality signallers must not merely suffer higher advertising costs than their higher quality competitors, but must suffer advertising costs that are higher *by some amount large enough to balance any fitness benefits accrued through signalling*.

Conversely, under the condition explored by Grafen, in which the beneficial fitness consequences of female preference are higher for higher quality signallers (i.e., $w_{23} > 0$), equation (7.4) reduces to,

$$w_{13} > -w_{23} * |\frac{w_1}{w_2}|.$$

It is equally plain that whilst, as Grafen maintains, ensuring that $w_{13} > 0$ is *sufficient* to ensure a solution to this inequality, it is not *necessary*. This inequality admits of solutions in which $w_{13} < 0$, i.e., non-handicap equilibria exist.

7.6 An Evolutionary Simulation Model of a Continuous Signalling Game

With the analytically derived results in hand the way is clear to design a simulation with which to examine the behaviour of the model under various conditions. Before such an evolutionary simulation model can be attempted, fitness functions which adequately capture the assumptions made during the above analysis must be defined for both signallers and receivers. These fitness functions will play the role of the discrete functions schematised in figure 7.1. Whereas the fitness formulae for the discrete simulation presented in section 7.3 followed trivially from the associated discrete model presented in section 7.2, merely requiring the selection of suitable ranges of values for costs and benefits, the construction of continuous fitness functions based upon Grafen's (1990b) model requires a little more thought. Particular attention will be paid to the ability of the constructed fitness functions to capture the assumptions made both by Grafen (1990b), and Godfray (1991). In order to allow this, quality will be defined as varying inversely with need.

In addition, schemes for representing a range of continuous signalling and response strategies must be defined to take the place of the eight discrete strategies depicted in table 7.1. The formulation of such schemes also requires some thought. They must be simple in order that the representation of strategies be amenable to manipulation by a genetic algorithm, yet they must also be able to capture an adequate range of signalling and responding behaviours.

7.6.1 Continuous Fitness Functions

After Grafen (1990b), observer fitness, w_O, may be calculated as

$$w_O = \frac{1}{1+|p-q|} \tag{7.5}$$

That is, observer fitness increases with the accuracy with which the observer response, p, approximates signaller quality, q.

Alternatively, in order to model the honest advertisement of need, in which observers are selected to respond positively to signallers with high need (e.g., Godfray, 1991), given some notion of need varying inversely with quality, a different observer fitness function must be constructed. For example,

$$w_O = \frac{1}{1+|p-n|} \tag{7.6}$$

specifies that observer fitness increases with the accuracy with which observers can predict need, n, where need varies inversely with quality, thus,

$$n = q_{max} - q + q_{min}$$

Specifying one of these two observer fitness functions is analogous to specifying one of the two pairs of inequalities which govern the fitness of observers within the discrete action-response game analysed above, i.e., inequalities (7.1), or (7.2). One observer fitness function confers one agenda on observers (e.g., to reward need), whilst the other confers the complementary agenda (e.g., to reward quality). Equation (7.5) will be described as an observer fitness function featuring a quality-rewarding agenda, and referred to as observer function (7.5), whereas equation (7.6) will be described as an observer fitness function featuring a need-rewarding agenda, and referred to as observer function (7.6).

Grafen (1990b) also constructed a specific function determining signaller fitness with which to demonstrate how his general model worked. This fitness function allows that increases in signaller quality (q) reduce the costs incurred in making an advertisement (a), and that increases in signaller quality increase the (positive) fitness consequences of female preference (p). The degree to which female preference influences signaller fitness is governed by a parameter, r.

$$w = p^r q^a$$

Grafen assumes that both q and p lie in the interval [0,1], and that both a and r are greater than or equal to unity.

As such, Grafen's function cannot accommodate the possibility that the fitness consequences of observer responses might vary with signaller quality independently from the manner in which the (negative) fitness consequences of advertising vary with signaller quality, nor that the (negative) fitness consequences for signallers of advertising might increase with signaller quality.

An alternative function must be constructed before an unconstrained exploration of the various possible signalling scenarios entertained within the literature can be undertaken.

$$w(a,p,q) = pq^R - aq^S \tag{7.7}$$

For this function, w, a, p, and q denote, as before, signaller fitness, level of advertising, degree of observer preference, and level of signaller quality, respectively, whilst R and S are indices which govern, respectively, the manner in which signaller quality mediates the positive effect of observer responses, and the manner in which signaller quality mediates the negative effect of signaller advertisement. The function is naturally understood as the sum of a positive benefit term and a negative cost term.

The first term of the fitness function, pq^R, connotes the benefit of signalling. The observer response, p, contributes positively to signaller fitness, but the manner in which it contributes may be sensitive to signaller quality. For scenarios in which $R = 0$, the fitness consequences of observer responses are independent of signaller quality. The maintenance of this equality is analogous to Hurd's (1995) stipulation that V is equal for all signallers, or to the maintenance of $w_{23} = 0$ within Grafen's (1990b) model.

For scenarios in which $R > 0$, the positive contributions of observer responses increase with signaller quality. Such scenarios are in line with the claim made by Grafen (1990b) that $w_{23} > 0$. For scenarios in which $R < 0$ the positive contributions of observer responses decrease with signaller quality. Such scenarios are analogous to the model constructed by Godfray (1991) to explore the honesty of begging nestlings.

The second term, aq^S, represents the cost of signalling. The signaller's level of advertisement, a, contributes negatively to signaller fitness, but the manner in which it contributes may be sensitive to signaller quality. For scenarios in which $S = 0$ the fitness consequences of advertising are independent of signaller quality. This condition is analogous to that stipulated within Godfray's (1991) model of begging nestlings.

For scenarios in which $S > 0$ the cost of advertising increases with q. Conversely, for scenarios in which $S < 0$ the cost of advertising decreases with q. This last class of scenarios is asserted by Zahavi (1975, 1977a) and Grafen (1990b) to be the only class admitting of honest signalling behaviour.

Thus, through manipulation of the signaller fitness function's two free parameters, R and S, and manipulation of the agenda of the observer fitness function, the continuous model can be made to explore the full range of scenarios so far discussed. It remains however to demonstrate that the function obeys the elementary strictures necessary for it to adequately model signalling.

First, advertising must be costly for signallers, and observer responses must be beneficial for them. These demands are satisfied by ensuring that the signs of the two partial derivatives of fitness with respect to advertisement and preference are correct.

As before, the condition that signals must be costly is maintained by the inequality, $w_1 < 0$,

$$\frac{\partial w}{\partial a} = -q^S.$$

That female preference is beneficial is similarly again maintained by the inequality, $w_2 > 0$,

$$\frac{\partial w}{\partial p} = q^R.$$

It is clear that the fitness function satisfies these two criteria. Grafen asserts that the second order partial derivatives, w_{13} and w_{23}, must be calculable. This is easily demonstrated.

$$w_{13} = \frac{\partial \frac{\partial w}{\partial a}}{\partial q} = -Sq^{S-1}$$

$$w_{23} = \frac{\partial \frac{\partial w}{\partial p}}{\partial q} = Rq^{R-1}$$

As we saw in the previous section, in order to derive the conditions for the existence of an honest signalling ESS in which observers reward high quality, we must derive the conditions under which Grafen's third condition, that "better males do better by advertising more" (Grafen, 1990b, p.520), is satisfied. Grafen formulated the condition thus,

$$\frac{\partial w/\partial a}{\partial w/\partial p} \text{ is strictly increasing in q},$$

This condition can be represented as

$$\frac{\partial(\frac{w_1}{w_2})}{\partial q} > 0.$$

Substituting for w,

$$\frac{\partial(\frac{-q^S}{q^R})}{\partial q} > 0,$$

and resolving the left side of the inequality yields

$$(R-S)q^{S-R-1} > 0,$$

which is satisfied exclusively by $R > S$. Thus we can expect honest signalling ESSs to exist for scenarios in which $R > S$, i.e., scenarios in which, naturally enough, the manner in which quality mediates the positive fitness consequences of female preference (R) should outweigh the manner in which quality mediates the negative fitness consequences of advertising (S).

It should be noted here that in order to model Godfray's (1991) account of begging nestlings using the scheme presented here, one of two approaches must be taken. The first approach is to use q to denote need rather than quality, in which case the ESS results derived above remain unchanged (this move is analogous to that discussed in section 7.4.1). Under this reading, the inequality, $R > S$, can now be interpreted as demanding that the manner in which *need* mediates the positive fitness consequences of female preference (R) should outweigh the manner in which *need* mediates the negative fitness consequences of advertising (S).

An alternative approach, discussed for the discrete action-response game within section 7.4.2, is to specify some notion of relatedness between the concepts of quality and need, whilst leaving the meaning of q unchanged. As in section 7.4.2, we might specify that quality, q, varies inversely with need, n, such that, for example,

$$n = q_{max} - q + q_{min}.$$

With such an understanding of the relationship between need and quality in place, the conditions under which the honest advertisement of n, is an ESS may be calculated as

$$\frac{\partial w/\partial a}{\partial w/\partial p} \text{ is strictly increasing in n},$$

that is,

$$\frac{\partial w/\partial a}{\partial w/\partial p} \text{ is strictly decreasing in q},$$

$$\frac{\partial(\frac{w_1}{w_2})}{\partial q} < 0,$$

$$\frac{\partial(\frac{-q^S}{q^R})}{\partial q} < 0,$$

$$(R-S)q^{S-R-1} < 0,$$

$$R < S.$$

As noted for the discrete action-response game in section 7.4.2, under such a reading of quality and need, no conditions exist in which a signal (e.g., a) may, in response to a change in observer agenda, be used to honestly advertise either need or quality (i.e., the intersection between the inequalities $R > S$, and $R < S$ is empty).

Godfray's result, that honest advertisement of need is an ESS under conditions in which signalling costs are unaffected by signaller need ($S = 0$), but signaller benefits increase with signaller need ($R < 0$), is clearly accommodated by the above result.

Thus a scheme of fitness functions has been presented which may be manipulated in order to capture the assumptions of various models within the literature. In addition, a clear prediction concerning the conditions under which honest advertisement is an ESS have been made. These conditions vary with the observer agenda. For scenarios in which observers are selected to respond positively to signallers of high *quality*, $R > S$ defines conditions under which the honest advertisement of quality is evolutionarily stable. In contrast, for scenarios in which observers are selected to respond positively to signallers of high need, $R < S$ defines conditions under which the honest advertisement of need is evolutionarily stable. These ESS conditions accommodate results presented within, for example, Grafen (1990b), and Godfray (1991), whilst allowing the existence of a broader class of honest signalling conditions than predicted under such models. This broader class of ESS conditions includes conditions in which Zahavi's handicap conditions do not have to be met (i.e., conditions in which $S \nless 0$).[5]

7.6.2 Continuous Signalling and Response Strategies

A population of signallers/receivers was distributed across a 25-by-25 grid. Each cell in the grid contained one signaller and one receiver. Each signaller was allocated an internal state, q, drawn at random from a uniform probability distribution in the range $[q_{min}, q_{max}]$. In addition, each signaller inherited a signalling strategy from its parent. A signalling strategy comprised two real values, $\{\theta, c\}$. An advertisement, a, was calculated as $q \tan\theta + c$. Advertisements of below zero were truncated to zero. Similarly each receiver inherited a response strategy from its parent. A response strategy comprised two real values, $\{\alpha, d\}$. Receiver response, p, was calculated as $a \tan\alpha + d$. Responses lying outside the range $[q_{min}, q_{max}]$ were truncated to their nearest extreme.

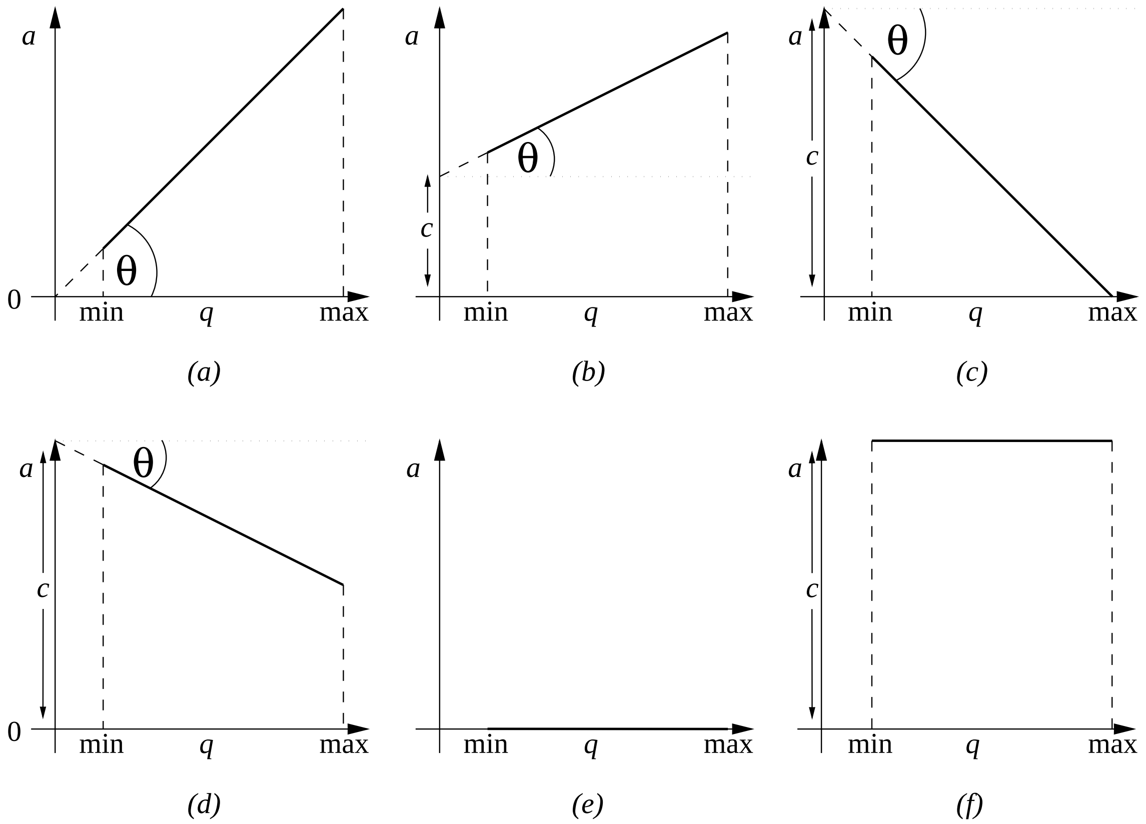
The honesty of such a signalling strategy cannot be ascertained through consideration of the strategy in isolation. Any of the signalling strategies depicted in figures 7.7a through 7.7d could take part in an honest signalling scenario since they each provide a unique advertisement for each possible value of signaller quality. Figures 7.8a through 7.8d depict response strategies which would successfully recover the value of q from advertisements made by signallers adopting the signalling strategies depicted in figures 7.7a through 7.7d respectively.

A population of signallers adopting any signalling strategy which maps every value of q onto a unique value of a partnered by a population of receivers adopting any response strategy which maps every value of a generated by the signalling population onto a unique value of p can be considered to be engaging in communication as defined by Enquist (1985). It is interesting to note that nonlinear strategies of the kind depicted in figure 7.8f do not satisfy these conditions as they either do not specify a unique advertisement for each value of internal state, or do not specify a unique response to each value of advertisement. However, they would not appear to necessarily result in non-communication (see section7.7).

Whether communication is deceitful or honest is contingent upon the manner in which the signalling and response strategies match up across the population. A population of signallers playing the strategy depicted in figure 7.7b partnered by a population of receivers playing the strategy depicted in figure 7.8b can be considered, in concert, to be taking part in an honest signalling

---

[5] N.b. The relationship between previous theoretical results, and the conclusions reached here are depicted in figure 7.20b below.



(a) *(b)* *(c)*

(d) *(e)* *(f)*

Figure 7.7: Examples of continuous signalling strategies mapping signaller quality, q, onto advertisement, a. Each strategy is defined by two parameters, θ and c, where $a = q\tan\theta + c$, with $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$, and c unbounded. Advertising level is truncated at zero, signaller quality ranges between q_{min} and q_{max}. Strategies *(a)*, $\{\theta = \frac{\pi}{4}, c = 0\}$, *(b)*, $\{\theta = \arctan(\frac{1}{2}), c = \frac{q_{max}-q_{min}}{2}\}$, *(c)*, $\{\theta = -\frac{\pi}{4}, c = q_{max}\}$, and *(d)*, $\{\theta = -\arctan(\frac{1}{2}), c = q_{max}\}$, each prescribe one to one mappings of $q \Rightarrow a$ (i.e., they prescribe advertisements which differ with signaller quality) and are thus informative (*sensu* Enquist, 1985). Strategies *(a)* and *(b)* differ from *(c)* and *(d)* in that they prescribe advertising functions which preserve, rather than invert, the rank order over q. Strategies *(e)* and *(f)* are examples of advertising strategies which prescribe a many to one mapping of $q \Rightarrow a$. Such strategies are not informative (*sensu* Enquist, 1985).

scenario since observer prediction error is minimised in such a situation. In such a population, a mutant signaller playing the alternative signalling strategy depicted in figure 7.7a must be considered to be cheating in that the quality of such a mutant would be consistently overestimated by receivers. The classification of such a signaller as a cheat must be made despite the fact that the particular signalling strategy employed by the mutant generates advertisements which are directly proportional to its internal state (see chapter 5 for discussion of issues related to the meanings of cheating and bluffing).

This scheme for the representation of signalling and response strategies compares favourably with alternative schemes proposed within similar models. For example, de Bourcier and Wheeler (1994, 1995; Wheeler & de Bourcier 1995) construct a model of aggressive signalling with which to explore the handicap principle, and propose that a signalling strategy can be represented as the

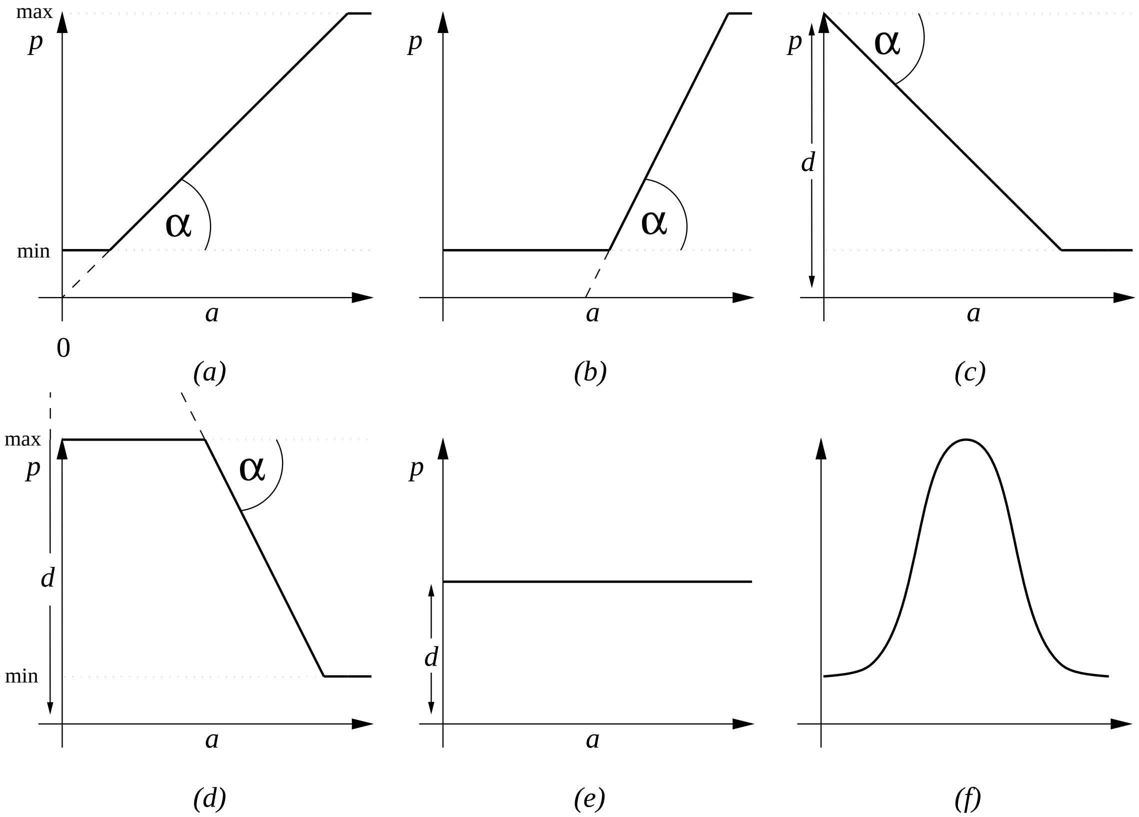


Figure 7.8: Examples of continuous response strategies mapping advertisement, a, onto receiver response, p. Each strategy is defined by two parameters, α and d, where $p = a\tan\alpha + d$, with $-\frac{\pi}{2} < \alpha < \frac{\pi}{2}$, and d unbounded. Receiver response is truncated to lie within the range $[q_{min}, q_{max}]$. Strategies *(a)*, $\{\alpha = \frac{\pi}{4}, d = 0\}$, *(b)*, $\{\alpha = \arctan(2), d = q_{min} - q_{max}\}$, *(c)*, $\{\alpha = -\frac{\pi}{4}, d = q_{max}\}$, and *(d)*, $\{\alpha = -\arctan(2), d = q_{min} + 2q_{max}\}$, successfully recover the value of q advertised under the strategies depicted in figures 7.7a through 7.7d. Strategy *(e)*, $\{\alpha = 0, d = \frac{q_{max} - q_{min}}{2}\}$, is an optimal response to any non-informative signalling strategy (e.g., a strategy in which $\theta = 0$ since it prescribes a single response, $p = \frac{q_{min} + q_{max}}{2}$, equal to the mean quality of a uniform distribution of signallers, $[q_{min}, q_{max}]$. Although the scheme used to represent strategies within the simulation admits a degree of non-linearity (due to the possibility of truncation at extreme values), the class of non-linearity exhibited by strategy *(f)* cannot be captured.

gradient, m, of an advertising function of the form $a = mq$. Under such a scheme, although signallers may employ different degrees of exaggeration, no signaller is able to signal more strongly when low quality than when high, and every signaller must make an advertisement of zero when of zero quality. This overly restricts the strategy space and consequently limits the evolutionary dynamics of their model.

7.6.3 Algorithm and Parameters

The fitnesses of receivers were calculated as per observer function (7.5), whilst that of signallers was calculated as per equation (7.7), each signaller interacting once with the receiver sharing its cell. Once each signaller and receiver had been assessed the whole population was updated synchronously. A parent from the previous generation was chosen for each offspring cell. The location of a potential parent was chosen through perturbing both the x and y grid co-ordinates of the offspring cell by independent values drawn from a normal probability distribution with standard deviation 1.75 and mean zero. Four potential parents were chosen for each offspring signaller. An offspring signaller inherited its signalling strategy from the fittest of these four. Similarly, an offspring receiver inherited its response strategy from the fittest of four receivers chosen from the previous generation in the same manner.

A mutation operator ensured that offspring sometimes inherited a strategy which differed from that of their parents. For both signallers and receivers each of the two values comprising their inherited strategy were independently exposed to the chance of mutations which occurred with probability 0.01. Mutations, when they occurred, consisted of perturbations drawn from a normal distribution with mean zero, and standard deviation 0.05. Mutated values which lay outside the legal range for the parameter they coded for were truncated to the nearest legal value for that parameter (see section 6.7.1 for a discussion of issues concerning such mutation operators).

For all simulations reported here, $q_{min} = 0.1$, $q_{max} = 5.0$, $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$, $-\frac{\pi}{2} < \alpha < \frac{\pi}{2}$, whilst both c and d were unbounded.

Populations were simulated for 1000 generations in this manner, during which time the signalling and response strategies present in the population were recorded.

The parameters R and S were varied across simulations but remained constant throughout each. Each of the 441 possible pairs of parameter values, $\{R,S\}$, drawn from the set {-2.0, -1.8, ... 1.8, 2.0}, was explored under each of three differing classes of initial condition. Each of the resulting 1323 (3 by 441) conditions were simulated 10 times. The pseudo-random number generator employed by the algorithm was itself seeded randomly for each simulation.

The first class of initial conditions consisted of a population of signallers sharing a signalling strategy which mapped q directly onto a, $\{\theta = \frac{\pi}{4}, c = 0\}$, and a population of receivers sharing a response strategy, $\{\alpha = \frac{\pi}{4}, d = 0\}$, which faithfully recovers values of q from signaller advertisements produced under the signalling strategy $\{\theta = \frac{\pi}{4}, c = 0\}$. This class of initial conditions will be termed ‘Honest’ since receivers are able to accurately predict signaller quality from signaller advertisements.

The second class of initial conditions consisted of a population of signallers, each with a signalling strategy generated by drawing a value for θ at random from a uniform distribution $[-\frac{\pi}{4}, \frac{\pi}{4}]$, and similarly drawing a value for c at random from a uniform distribution $[-q_{max}, q_{max}]$, and a population of receivers, each with a response strategy generated by drawing a value for α at random from a uniform distribution $[-\frac{\pi}{4}, \frac{\pi}{4}]$, and similarly drawing a value for d at random from a uniform distribution $[-q_{max}, q_{max}]$. This class of initial conditions will be termed ‘Random’ since signallers’ strategies and receivers’ strategies are unrelated, and implement a wide range of mappings.

The third class of initial conditions consisted of a population of signallers sharing a signalling strategy which mapped any value of q onto 0, i.e., $\{\theta = 0, c = 0\}$, and a population of receivers sharing a response strategy which mapped any advertisement onto 0, i.e., $\{\alpha = 0, d = 0\}$. This class of initial conditions will be termed ‘Cynical’ since signallers never make advertisements, whilst receivers never make responses.

7.6.4 Results

Results were consistent with the predictions arrived at through the analysis presented above. Two measures of performance were utilised in assessing the degree of honesty within a population. Both measures were derived from population summary statistics calculated for a particular generation. First, the average signalling strategy and response strategy were calculated. This was achieved simply by taking the population mean values of θ, c, α, and d.

From the mean signalling strategy, $\{\overline{\theta},\overline{c}\}$, the mean strategy signal range, $\overline{r}$, was calculated as $(q_{max} - q_{min})\tan\overline{\theta}$. The mean strategy response error, $\overline{e}$, was calculated as the mean difference between between signaller quality, q, and receiver response, p, for signallers using signalling strategy $\{\overline{\theta},\overline{c}\}$, and receivers using the mean response strategy $\{\overline{\alpha},\overline{d}\}$, calculated for q ranging from q_{min} to q_{max}.

Thus, a signalling population for which $\overline{r} \approx 0$ is one in which signaller quality has little effect upon advertisement, whereas a signalling population for which $\overline{r} > 0$ is one in which advertisements typically increase with signaller quality, whilst, conversely, a population in which $\overline{r} < 0$ is one in which advertisements typically decrease with signaller quality.

Similarly, a receiver population in which $\overline{e} \approx 0$ is one in which receivers can accurately predict signaller quality on the basis of signaller advertisement, whereas a population in which $\overline{e} \approx \frac{q_{max}+q_{min}}{4}$ is one in which receivers are, on average, predicting signaller quality at chance levels, whilst a receiver population in which $\overline{e} > \frac{q_{max}+q_{min}}{4}$ is one in which, on average, receivers are predicting signaller quality at worse than chance levels.

Throughout this section, the phrase ‘performing at chance levels’ will be used to refer to a receiver strategy which consistently predicts that a signaller has $q = \frac{q_{max}+q_{min}}{2}$, irrespective of any advertisement a signaller might make. Since, within the signaller population, quality is described by a flat probability distribution, $[q_{min}, q_{max}]$, such a strategy will result in a mean response error of $\frac{q_{max}+q_{min}}{4}$. This receiver strategy is thus an optimal response to any uninformative signalling strategy (e.g., strategies depicted in figures 7.7e and 7.7f).

These metrics allow the evolutionary dynamics of the simulation model to be characterised in an efficient manner. From prior analysis it is predicted that, simulations for which $R > S$, will evolve such that $\overline{e} \approx 0$, and $\overline{r} > 0$ (i.e., honest signalling), whereas, simulations for which $R < S$, will evolve such that $\overline{e} \approx \frac{q_{max}+q_{min}}{4}$, and $\overline{r} \approx 0$ (i.e., non-signalling). These predictions would, of course, be reversed if the fitnesses of receivers were calculated as per observer function (7.6), rather than observer function (7.5). However, although these predictions characterise the general style of equilibria expected, they take no account of any initial conditions in which signaller and receiver population might find themselves.

Honest Initial Conditions

From Honest initial conditions the equality $R = S$ divided the parameter space into two areas (see figures 7.9, 7.10, and 7.11). The area defined by $R > S$ contained signallers which made advertisements which increased with signaller quality, and receivers which were able to accurately recover signaller quality from such advertisements (i.e., honest signalling obtained under these conditions). In contrast, the area defined by $R < S$ contained signallers which made advertisements which did not differ with signaller quality, and, as a result, receivers which were unable to accurately recover signaller quality from signaller advertisements (i.e., non-signalling obtained under these conditions).

Mean signal range, $\bar{r}$, increased with $R - S$. For scenarios in which the difference between R and S is small, the range of signals is also small, however, for scenarios in which R far outstrips S, signals given by high quality signallers are orders of magnitude higher than those given by low quality signallers.

Mean response error for simulations in which $R < S$ varied considerably. Although the fact that signallers within these populations did not provide informative advertisements might lead one to expect that receivers would simply provide an average response to any signal, and thus achieve a mean response error of $\frac{q_{max}+q_{min}}{4}$, receivers under these conditions frequently performed at levels significantly worse than chance. This suggests that, from Honest initial conditions, receivers found it hard to evolve the strategy $\{\theta = 0, c = \frac{q_{max}+q_{min}}{2}\}$, which is an optimal response to uninformative signalling strategies. This phenomenon can be explained as an artefact of the simulation design through considering the coupled evolutionary dynamics of the signalling and receiving populations (see figures 7.12 and 7.13).

When $R < S$, honest signalling is unstable. As a result, over evolutionary time, signallers reduce the level of their advertising, i.e., θ decreases. Over the same evolutionary period, in order to successfully capture q from these decreasing advertisements, receivers adopt increasingly amplificatory response strategies, i.e., α increases. These complementary trends are depicted between generations 0 – 810 in figure 7.12, and generations 0 – 470 in figure 7.13.

Eventually this trend will lead to signalling strategies which produce signals with a mean signal range of zero, i.e., uninformative signals. Once signallers have reduced their advertisements to zero, amplificatory receiver strategies will fail to recover signallers' quality, constantly producing an observer response of roughly zero. Since receiver response, p, is constrained to lie between q_{min} and q_{max}, and, for the simulations considered here, $0 < q_{min}$, these response strategies result in a mean response error of $\bar{e} = \frac{q_{max}+q_{min}}{2}$, twice that which would result from performance at chance levels.

Although there is selective pressure on receivers to evolve strategies which minimise response error, this pressure is not effective, since α is free to vary without any impact on receiver fitness, and d is likewise free to vary without any impact on receiver fitness providing $d < q_{min}$, which it typically is.

Under these conditions, response strategy parameter values will drift at random, (generation 810 onwards in figure 7.12; generations 470 – 570 in figure 7.13), until $d > q_{min}$ (generation 570 in figure 7.13). Once $d > q_{min}$ selective pressure will decrease response error, through increasing d (generations 570 onwards in figure 7.13), until it reaches the optimum under conditions in which signallers provide uninformative advertisements.

Thus Honest initial conditions retard evolution toward the optimal receiving strategy under conditions in which honest signalling is not an ESS. However, the manner in which this retardation takes place is complex, and allows us to predict that in situations in which $q_{min} \gg 0$, receiver performance at chance levels may only be achieved after very long periods of evolutionary drift, during which time mean response error will be at twice that predicted for non-signalling equilibria.

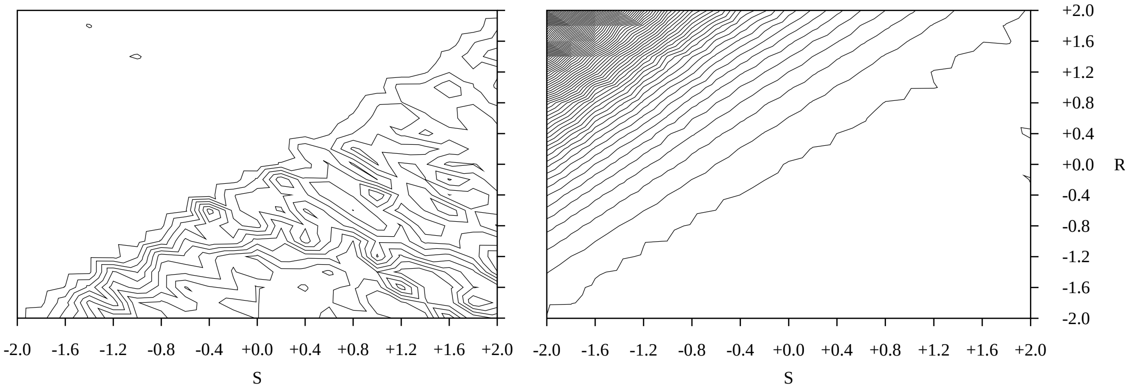


Figure 7.9: Mean response error (left) and signal range (right) after 1000 generations, averaged for 10 simulation runs from Honest initial conditions. The behaviour exhibited by the simulation model under Honest initial conditions can be divided into two classes by the equality $R = S$. For $R < S$, mean signal range is very low, whereas mean response error is high (at times above the error that would result from blind guessing). For $R > S$, mean signal range is positive, increasing with the difference between R and S, whilst mean response error is approximately zero.

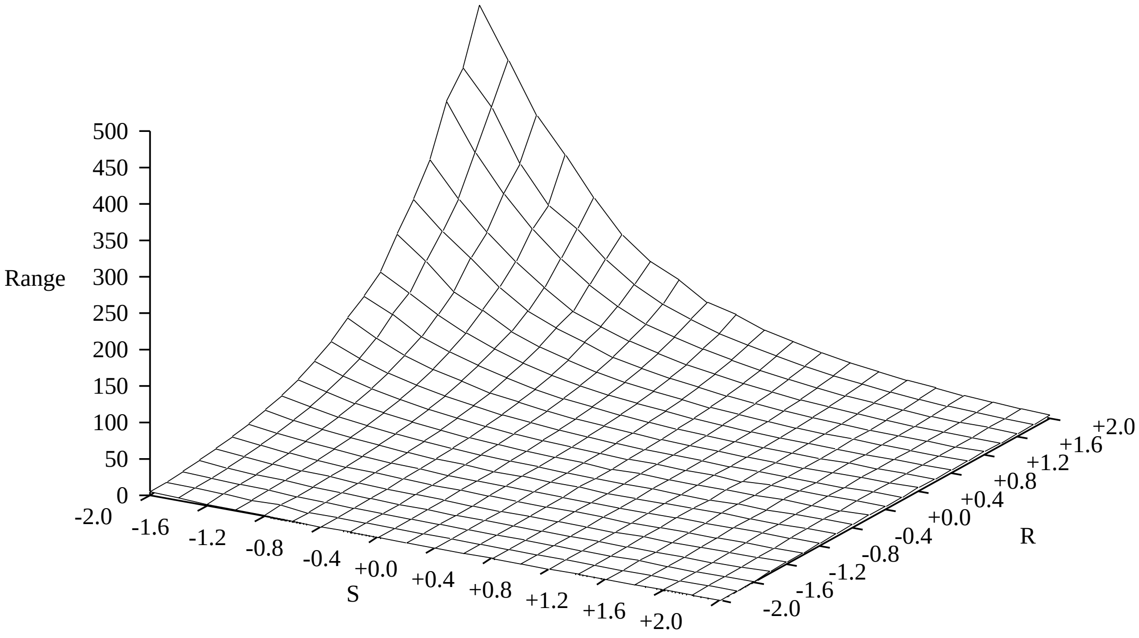


Figure 7.10: Mean signal range after 1000 generations, averaged for 10 simulation runs from Honest initial conditions. See figure 7.9 and text for details.

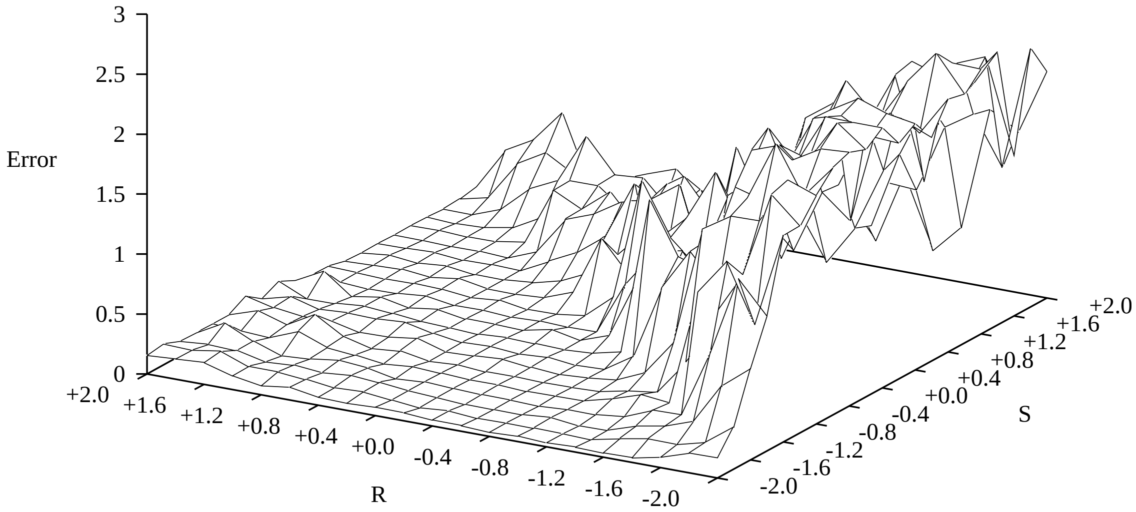


Figure 7.11: Mean response error after 1000 generations, averaged for 10 simulation runs from Honest initial conditions. See figure 7.9 and text for details.

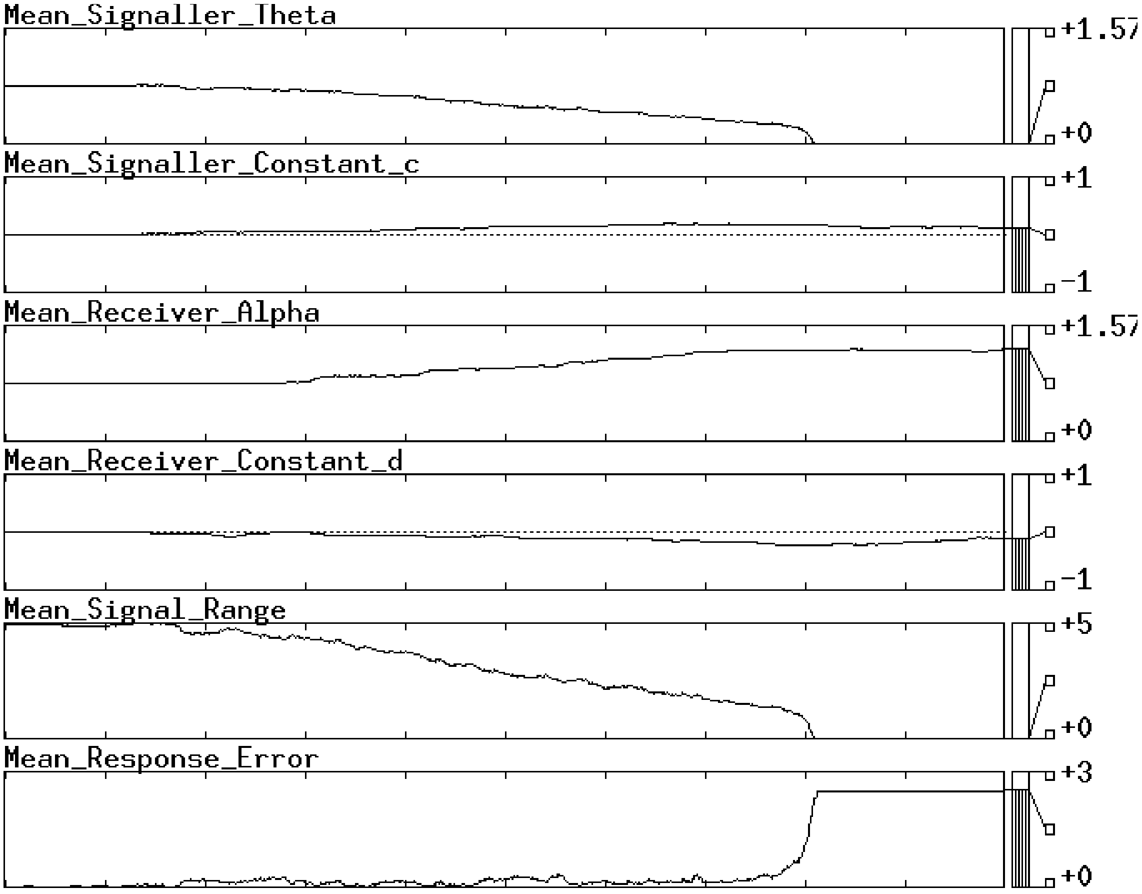


Figure 7.12: A trajectory through strategy space produced by a population of signallers and receivers evolved from Honest initial conditions, with $R = -2.0$, and $S = +1.8$. Each plot depicts the population mean value of a trait or statistic calculated for each of 1000 generations. Receivers evolve to a sub-optimal response strategy (generation 810) from which drift does not dislodge them. See text for details.

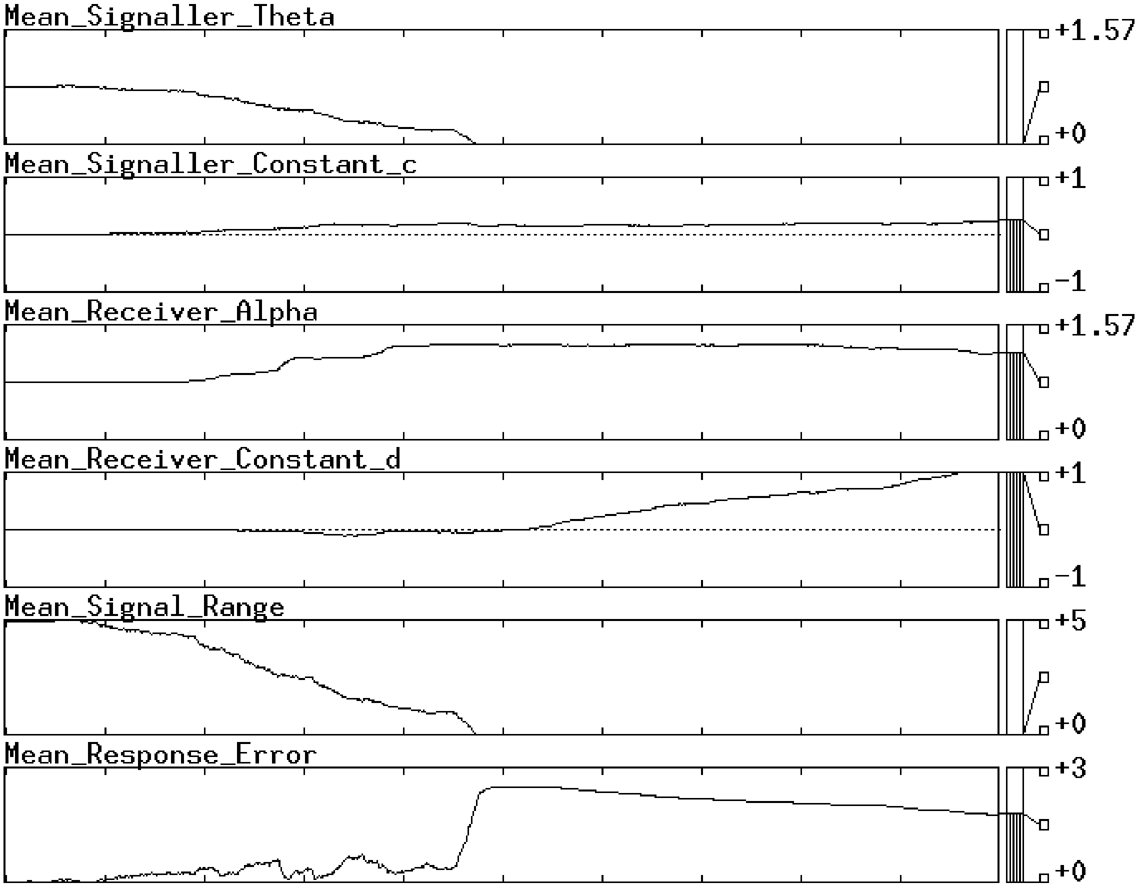


Figure 7.13: A trajectory through strategy space produced by a population of signallers and receivers evolved from Honest initial conditions, with $R = -2.0$, and $S = +1.8$. Each plot depicts the population mean value of a trait or statistic calculated for each of 1000 generations. Receivers evolve to a sub-optimal response strategy (generation 470) from which genetic drift eventually dislodges them (generation 570). See text for details.

Random Initial Conditions

From Random initial conditions, the behaviour of the model is similar to that resulting from Honest initial conditions (see figures 7.14, 7.15, and 7.16). Mean signal range is very low in the area of parameter space in which honest signalling equilibria are not predicted to exist (i.e., conditions in which $R < S$). In addition, simulations within this area of parameter space result in mean response errors approximately equal to performance at chance. In contrast, within the area of parameter space predicted to contain honest signalling equilibria, as for simulations from Honest initial conditions, mean signal range is positive, and increases with $R - S$, whilst mean response error is better than performance at chance.

However, the average mean signal range is lower than that resulting from initial conditions, and average mean response error is higher. These results are due to the simulation's inability to consistently reach honest signalling equilibria from random initial conditions. Failure to discover an honest signalling equilibrium results in simulations reaching non-signalling equilibria at which no signalling occurs and receivers perform at chance levels. The frequency with which the simu-

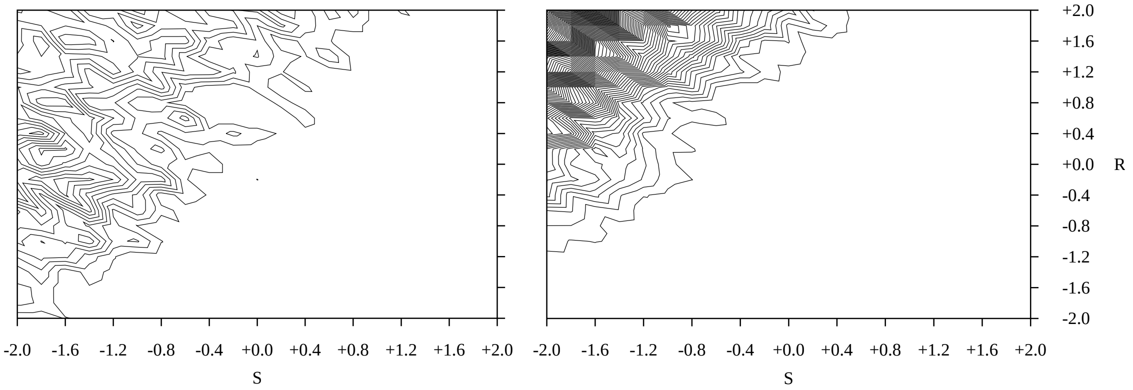


Figure 7.14: Mean response error (left) and signal range (right) after 1000 generations, averaged for 10 simulation runs from Random initial conditions. The behaviour exhibited by the simulation model under Random initial conditions can be divided into two classes by the equality $R = S$. For $R < S$, mean signal range is very low, whereas mean response error is approximately equal to that resulting from performance at chance. For $R > S$, mean signal range is positive, increasing with the difference between R and S (but at a far slower rate than that resulting from Honest initial conditions). Simulations in which $R > S$ do not always reach an honest signalling equilibrium, but instead reach a stable non-signalling equilibrium. The relative frequency with which simulations discover honest signalling equilibria increases with $R-S$, resulting in a decrease in mean response error as R increases and S decreases.

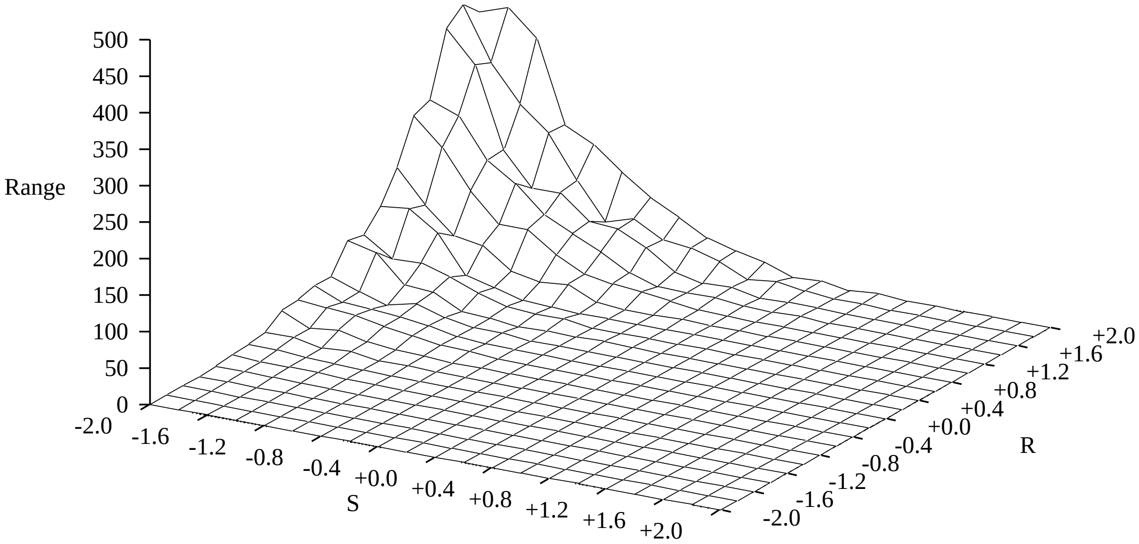


Figure 7.15: Mean signal range after 1000 generations, averaged for 10 simulation runs from Random initial conditions. See figure 7.14 and text for details.

lation arrives at honest signalling equilibria increases with $R-S$. This is reflected in the variation in average values for $\bar{r}$ and $\bar{e}$ across the parameter space.

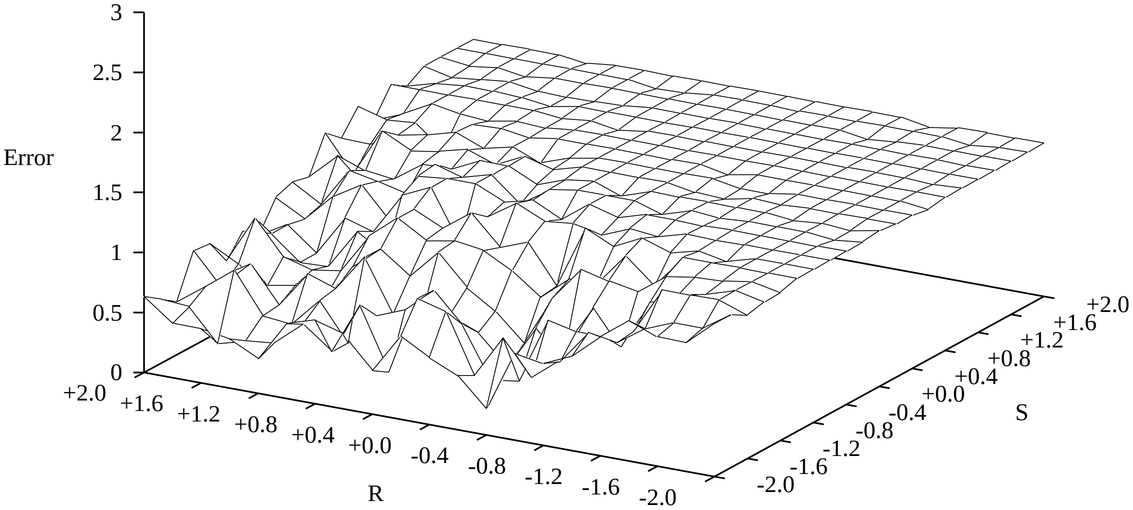


Figure 7.16: Mean response error after 1000 generations, averaged for 10 simulation runs from Random initial conditions. See figure 7.14 and text for details.

Cynical Initial Conditions

From Cynical initial conditions, the behaviour of the model is similar to that resulting from Random initial conditions (see figures 7.17, 7.18, and 7.19). For simulations in which $R < S$, (i.e., simulations predicted to admit of no honest signalling equilibria) behaviour is comparable to that resulting from Random initial conditions. For simulations which are predicted to admit of honest signalling equilibria, the rate at which such equilibria are achieved is even lower than that resulting from Random initial conditions. As with simulations simulations run from Random initial conditions, the frequency of attainment of honest signalling equilibria increases with $R - S$ for simulations run from Cynical initial conditions.

7.6.5 Summary

Grafen's (1990b) continuous model of signal evolution has been extended in a manner analogous to that carried out for Hurd's discrete action-response game. It has been demonstrated analytically that non-handicap equilibria exist for this extended model (i.e., equilibria exist for scenarios in which $w_{13} < 0$). An evolutionary simulation model constructed to explore the dynamic character of the model has supported this analytic result (i.e., honest signalling equilibria exist for scenarios in which $S > 0$), and demonstrated that, unlike the discrete model explored previously, the continuous game admits of no conventional signalling equilibria, but settles eventually on either an honest signalling equilibria, or a non-signalling equilibria (although from Honest initial conditions, the time taken to achieve non-signalling equilibria may be very great).

It is not surprising that oscillatory, conventional-signalling trajectories do not exist for this continuous game. Whilst, for a game involving a discrete strategy space, populations may oscillate between two adjacent strategies, each of which is unstable with respect to the other, a game involving a continuous strategy space will feature an infinitude of intervening strategies, one or more of which may offer a stable compromise.

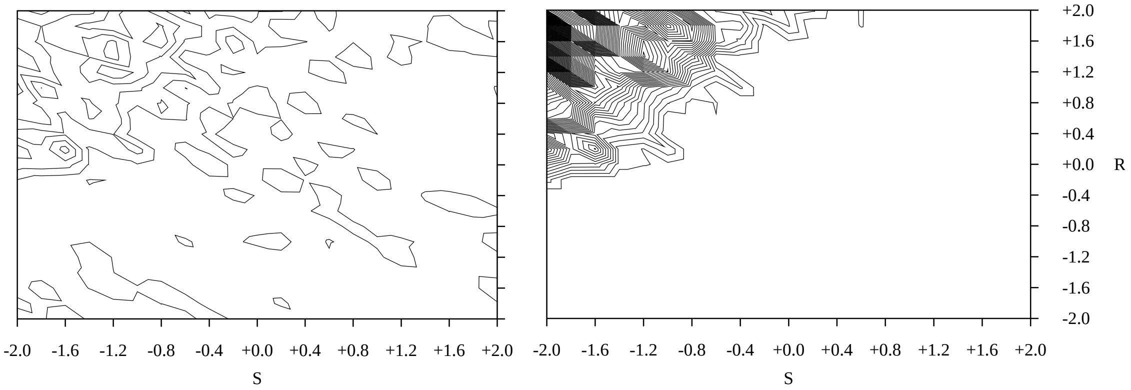


Figure 7.17: Mean response error (left) and signal range (right) after 1000 generations, for 10 simulation runs from Cynical initial conditions. The behaviour exhibited by the simulation model under Cynical initial conditions can be divided into two classes by the equality $R = S$. For $R < S$, mean signal range is very low, whereas mean response error is approximately equal to that resulting from performance at chance. For $R > S$, mean signal range is positive, increasing with the difference between R and S (but at a far slower rate than that resulting from Random initial conditions, and an even slower rate than that resulting from Honest initial conditions). Simulations in which $R > S$ rarely reach an honest signalling equilibrium, but instead reach a stable non-signalling equilibrium. The relative frequency with which simulations discover honest signalling equilibria increases with $R - S$, resulting in a decrease in mean response error as R increases and S decreases.

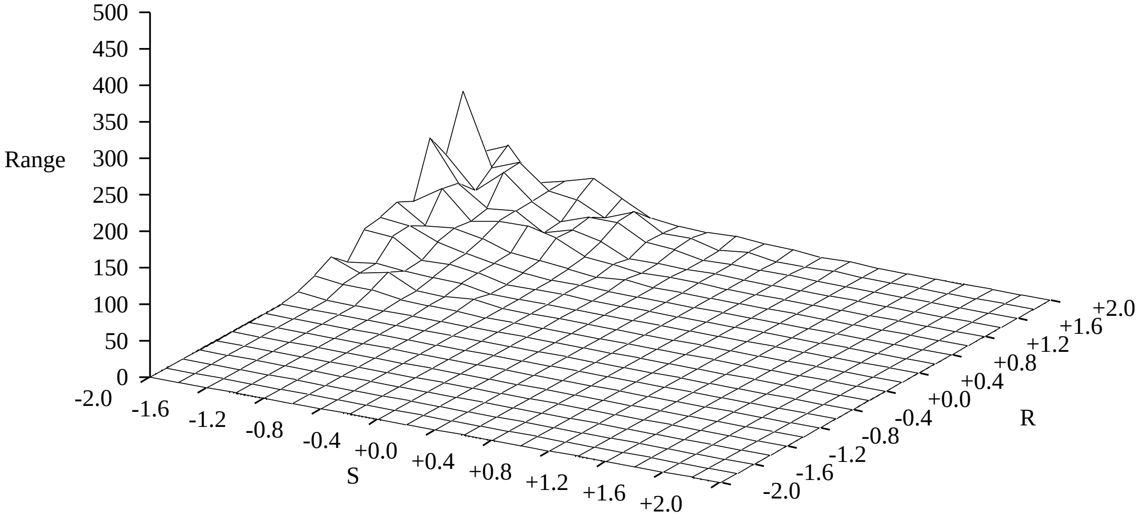


Figure 7.18: Mean signal range after 1000 generations, for 10 simulation runs from Cynical initial conditions. See figure 7.17 and text for details.

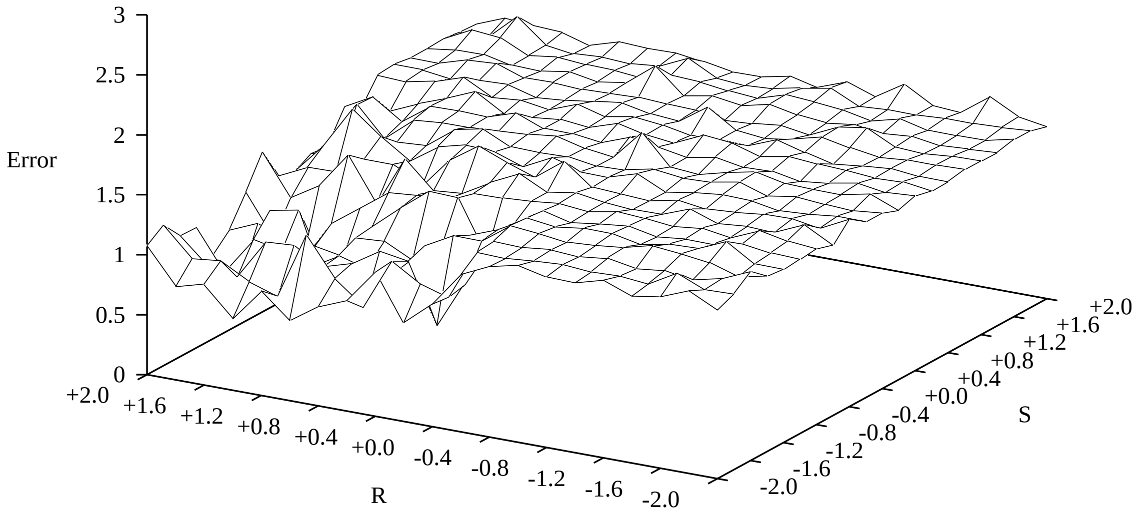


Figure 7.19: Mean response error after 1000 generations, for 10 simulation runs from Cynical initial conditions. See figure 7.17 and text for details.

7.7 Discussion

In this section, the findings reported in this chapter will be compared and contrasted with those reported within the studies which this chapter has extended upon. These previous results are easily accommodated by those presented within this chapter, which themselves provide a general formulation of the conditions governing the existence of what have been termed ‘handicap’ signalling scenarios.

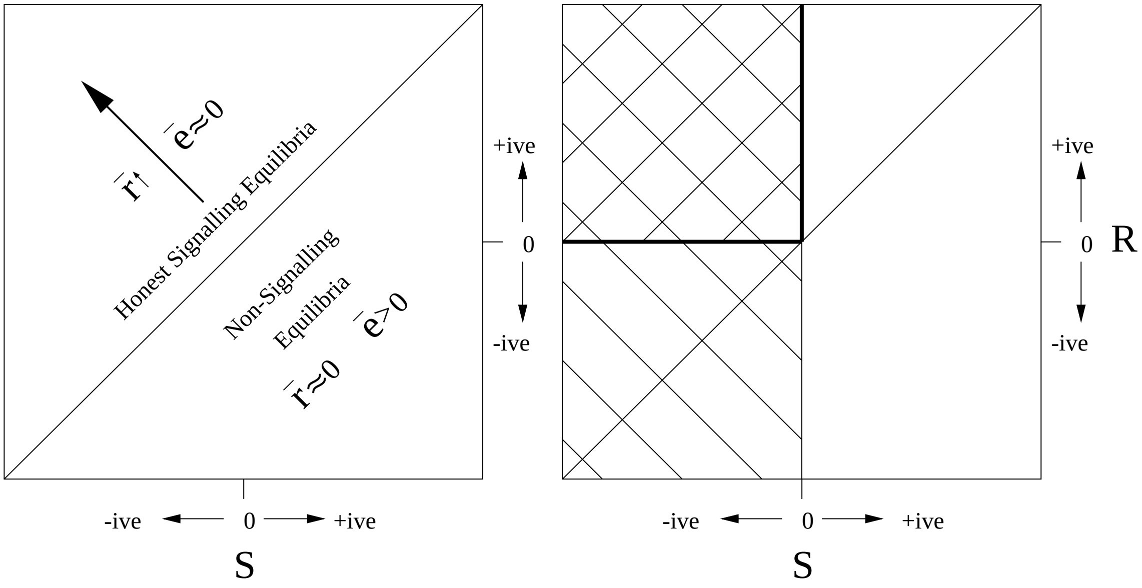
Once this reconciliation of previous results has been described, a reconciliation of the positions which lead to their presentation will be attempted. The handicap principle will be assessed in three regards. The first issue discussed will be the various interpretations of the relationship between costs, benefits, and fitness which appear to motivate models of the handicap principle. The second issue to be addressed will be the implications of the results presented here for the supposed generality of the handicap principle. Thirdly, some consideration will be paid to the notion that handicap signalling scenarios should in fact be considered to be co-operative signalling scenarios in the sense of co-operative signalling introduced in chapter 5.

Finally, some possible extensions to the work carried out here will be outlined. Specifically, attention will be drawn to the relatively under-developed role of the observer fitness function, the absence of noise from the model, and the non-heritability of the advertised trait.

7.7.1 Reconciliation of Results

Figure 7.20a depicts the broad conclusion suggested by the continuous model of signalling explored in sections 7.5 and 7.6. Honest signalling is stable for scenarios in which the manner in which the advertised trait mediates the influence of signalling benefits on signaller fitness outweighs the manner in which the advertised trait mediates the influence of signalling costs on signaller fitness, i.e., the net cost of signalling (the cost of signalling offset by the benefit of an accurate observer response) increases monotonically with the advertised trait.

This result is captured graphically in figure 7.20a by dividing the space of possible signalling



(a) (b)

Figure 7.20: Showing *(a)* the conditions under which honest signalling obtain for the continuous model described in sections 7.5 and 7.6, and *(b)* the conditions predicted to admit of honest signalling equilibria by Zahavi (1975, 1977a, 1977b; diagonal hatching defined by $S < 0$), Grafen (1990b; diagonal hatching defined by $S < 0$ and $R \geq 0$), Godfray (1991; bold vertical line defined by $R > 0$ and $S = 0$), and Hurd (1995; bold horizontal line defined by $S < 0$ and $R = 0$). (N.b. In order to depict Godfray's (1991) result concerning advertisement of need on the same diagram as other results concerning advertisement of quality, we must assume that increasing q represents increasing need for the former, but increasing quality for the latter.)

scenarios into two halves, separated by a diagonal line along which the influence of the advertised trait on costs is exactly balanced by its influence on benefits (i.e., $R = S$). Above this line (i.e, for $R > S$) honest signalling equilibria obtain; below it (i.e., for $R < S$) no such honest signalling equilibria exist.

In figure 7.20b this graphical device is used to locate previous theoretical results. For example, Zahavi's (1975, 1977a, 1977b) claim that honest signalling may only exist for scenarios in which the costs of signalling increase with the trait being advertised may be represented by the area satisfying the inequality $S < 0$. It is plain from the diagram that this inequality is neither necessary, nor sufficient for the existence of honest signalling equilibria. Grafen's (1990b) contention that, given that signaller benefits are either unaffected by the advertised trait ($R = 0$), or increase with the advertised trait ($R > 0$), in order that signalling be honest, signalling costs must decrease with the advertised trait ($S > 0$) is shown to be correct. But the space of possible signalling scenarios defined by Grafen's conditions does not exhaustively account for all honest signalling.

Results in which the negative fitness consequences of signal costs are assumed to be independent of the trait being advertised (i.e., models for which $S = 0$) have often concluded that, in order for such signalling to be honest, the positive fitness consequences (for signallers) of observer behaviour must increase with the signaller's advertised trait (i.e., $R > 0$). Such models are typically those in which the advertised trait is taken to be signaller need (e.g., Godfray, 1991; Maynard Smith, 1991). They make a claim which can be recast as asserting that honest signalling may be stable for signalling systems which lie along the bold vertical line in figure 7.20b.

Similarly, results in which the positive fitness consequences of the benefits accrued by signallers are assumed to be independent of the trait being advertised (i.e., models for which $R = 0$) have often concluded that, in order for such signalling to be honest, the negative fitness consequences (for signallers) of signal cost must increase with the signaller's advertised trait (i.e., $S < 0$). Such models are typically models in which the advertised trait is taken to be signaller quality (e.g., Hurd, 1995), and make a claim which can be rephrased as asserting that honest signalling may be stable for signalling systems which lie along the bold horizontal line in figure 7.20b.

7.7.2 Reconciliation of Positions

This chapter opened with a description of two complementary arguments which each result from Zahavi's handicap signalling notion. The first argument suggested that honest advertisement of quality might be under-written by differential signaller costs. The second argument suggested that the honest advertisement of need might be under-written by differential signaller benefits. The results of the models constructed within this chapter demonstrate that the honest advertisement of either quality, *or* need, may each be stabilised by differential costs, *or* differential benefits. This result is due to the fact that the terms 'cost' and 'benefit' may each be cashed out in the same currency — fitness. Costs are merely negative increments to fitness, whereas benefits are positive increments to fitness.

However, at a less abstract level of description, costs and benefits may come in many different forms. For example, negative fitness consequences may arise as a result of energetic costs, costs in terms of risk of predation, parasitism, or infection, costs of missing a high-quality mating opportunity, costs of mating with a sub-optimal mate, etc. Although each of these costs has negative fitness consequences, the character of these negative fitness consequences may differ radically across these different forms of cost.

Similarly there are benefits to be gained from obtaining a copulation, obtaining a food resource, obtaining a territory, obtaining an opponent's surrender, etc. Again, although each of these benefits has positive fitness consequences, the character of these positive fitness consequences may not be uniform across these different forms of benefit.

Within evolutionary models, the manner in which costs and benefits influence fitness is formalised identically. Costs, whatever their nature influence fitness negatively, whereas benefits, whatever their nature influence fitness positively. Furthermore, within a game theoretic model, for example, the degree of this influence on fitness is typically considered to be constant across the range of strategies which may be employed within the game. For example, in the Hawk-Dove game (e.g., Maynard Smith, 1982) the positive fitness consequences derived from obtaining a food resource is considered to be the same for both Dove and Hawk. Similarly, the negative fitness consequences of engaging in combat are considered to be equal for both Dove and Hawk.

However, theorists constructing models of handicap signalling are faced with a decision concerning the manner in which the influence of signalling costs (and signalling benefits) upon fitness is to *vary* with the trait which signallers are advertising. For example, how does the effect of signal production cost vary with signaller need? What will interest us here are the different decisions which may be made regarding these aspects of handicap modelling, and the possible motivations which underly them.

Consider the example of a begging nestling which is signalling in an attempt to solicit parental resources. We will assume that the trait of interest to parents is a chicks hunger, and that nestlings of high quality experience low need (hunger), and vice versa, i.e., that quality is inversely proportional to need. For this scenario Godfray (1991) models the cost of signalling as equal across all signallers. Grafen (1990b), on the other hand, models cost as decreasing with signaller quality. Godfray (1991) models the benefit of soliciting a particular parental resource as increasing with need, whereas Grafen (1990b) models such benefit as *decreasing* with signaller need, or independent of signaller need.

A second example, also addressed by Grafen (1990b) involves an interloper making a signal of aggressive intent to an observing harem holder. Grafen asserts that in such a situation, the costs of signalling decrease with the increasing quality of a signaller. He further claims that the benefits (for the signaller) of a retreat response by an observer increase with the quality of a signaller. In contrast, Adams and Mesterton-Gibbons (1995) suggest that in such situations, the benefit of eliciting a retreat response might *decrease* with increasing signaller quality. They reason that “strong animals can win many conflicts without threatening (i.e., by direct fighting), while weak animals cannot. Furthermore, weak animals have more to gain by avoiding direct fights since they are less able to defend against injury.” (p.406).

It is clear from these two examples, that the authors of these models have made radically opposed assumptions with respect to the relationship between costs, benefits, and fitness. In contrast, the models presented within this chapter make no assumptions concerning the manner in which costs and benefits influence fitness, save that costs are a negative influence, whilst benefits are a positive influence. As a result of this neutrality, a degree of generality has been gained.

Absolute vs. Relative Costs

One manner in which the assumptions concerning the relationship between costs, benefits and fitness adopted by a theorist may affect the construction of her handicap model, and consequently the results obtained from such a model, can be captured by opposing an understanding of fitness as being influenced by the *relative* cost (or benefit) of a behaviour, with the understanding of fitness as being influenced by the *absolute* cost (or benefit) of a behaviour. A difference of this sort appears to underly the conflict between the models of Godfray (1991), Grafen (1990b), and Adams and Mesterton-Gibbons (1995) outlined above.

Consider a line of reasoning which might support the claim that Zahavi’s second handicap criterion (that poor quality signallers must pay more for a certain signal than their higher quality competitors) is true of natural signallers. “Poor quality signallers”, the reasoning runs, “pay higher signalling costs because, in proportion to their reserves, the energy expenditure, time expenditure etc., required for any signal is higher for poor quality signallers than for those of higher quality”. If this argument holds then $S < 0$.

However, this line of reasoning has a corollary in the claim that “Poor quality individuals gain more from a particular observer response than their higher quality competitors because any resource gain would be greater proportionally for poor quality signallers than for those of higher quality”. If this argument holds then $R < 0$.

These arguments rely on what I shall call a *relative* reading of the relationship between cost/benefit and fitness. Under this reading, although two signallers of differing quality use identical amounts of energy to produce a signal, the fitness consequences of making that signal differ as a result of the *relative* cost of signal production. From the perspective of a low quality signaller, the signal is *relatively* expensive, whereas from the perspective of a high quality signaller, it is *relatively* cheap. By relative I am referring to the energetic demands of signal production when compared to the signallers’ energy resources. Such a reading allows one to construct the corollary above.

This relativistic perspective on the relationship between cost and fitness (and therefore also between benefit and fitness) cannot justify the existence of a handicap unaided, as any claim to the effect that $S < 0$ based on the *relative* costs of signalling is attended by a complementary claim to the effect that $R < 0$ based on the relative benefits of observer responses. Therefore, in order to demonstrate that $R > S$, some further argument must be produced.

However, Zahavi’s exposition of the second handicap criterion sometimes seems to promote a more *absolute* account of signal costs. He claims that “it is reasonable to assume that high quality phenotypes and experienced individuals pay less for the cost of the same sized handicaps than low quality phenotypes” (1977, p. 604). The thought here, perhaps, is that the superior skills, metabolism, morphology, etc., of high quality phenotypes might just make signalling easier. This would result in a situation in which the absolute energetic expenditure required to make the same signal differs between signallers of differing quality. This absolute reading does not licence a corollary of the kind outlined above. In contrast, the benefit of an observer response might be considered to be best utilised by the same high quality individuals which find it easier to produce signals. For example, a particular worm might have a particular calorific value which could be best exploited by the metabolism of a large, fit, chick.

Such reasoning would support Grafen’s (1990b) contention that “the fitness gained by a marginal improvement in the parent’s assessment of a chick is at least as great for big as for small chicks” (p. 527). However, a more ‘relative’ reading of signalling costs/benefits seems to motivate Godfray’s (1991) model of offspring begging signals. He (directly reversing Grafen’s assumption) assumes that “the benefits of [solicited parental] resources increase with [offspring] need”.

The tension between these relative and absolute readings of the manner in which both cost and benefit influence fitness is not irreconcilable. One may maintain that there are both absolute *and* relative components to the fitness contributions of cost and benefit. The difference between relative and absolute interpretations of the relationship between, on the one hand, the costs and benefits of a behaviour and, on the other, the fitness implications of the behaviour are depicted in figure 7.21.

7.7.3 Handicap Signalling and Co-operative Signalling

It was suggested in chapter 5 that for a system to be properly regarded as a signalling system, one must be able to identify a producing system, and a consuming system. The evolutionary function of the producing system must be to effect a signal, the mapping-rule of which adapts a consumer to some state of affairs. The evolutionary function of the consuming system must be to respond to the signal in some manner which is adapted to the existence of the state of affairs to which the

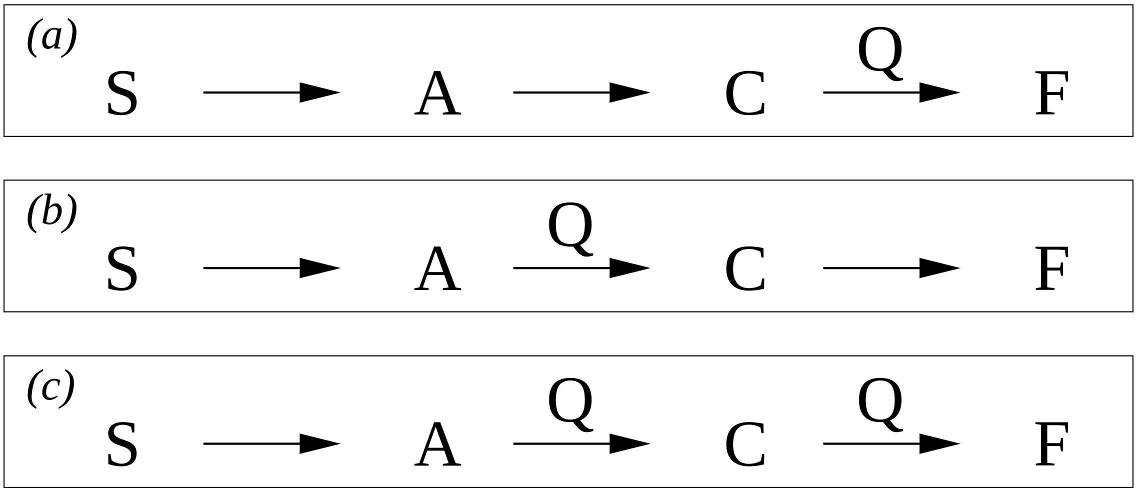


Figure 7.21: A signaller, *S*, of quality, *Q*, makes an advertisement, *A*, which leads to a raw cost, *C*, in terms of energy, risk, etc. This cost results in a change in fitness, F. *(a)* A relative reading of cost — *A* leads to *C* independent of *Q*, whilst the manner in which *C* influences F is mediated by *Q*. *(b)* An absolute reading of cost — the manner in which *A* leads to *C* is mediated by *Q*, whilst *C* influences F independently of *Q*. *(c)* Both — *Q* mediates the manner in which *A* leads to *C and* the manner in which *C* influences F.

signal's mapping-rule maps it.

As a result of this suggestion, the notion that full-blooded signalling necessarily involves both a producer and consumer which have coincident evolutionary agendas was mooted. Co-operative signalling systems were identified as paradigmatic examples of such producers and consumers. An example was made of bee communication which involves a dancer and audience with completely coincident agendas as a result of their reproductive potential being channeled through the same individual. Here the notion that handicap signalling may be similarly classed as co-operative, and thus full-blooded, will be explored.

The conditions for the stability of handicap signalling, derived within this chapter, stipulate that in order for handicap signalling to be honest the net cost of a signal must increase monotonically with the trait being advertised. In such a situation, a population of signallers ranked according to this trait will simultaneously be ranked according to their ability to make a signal of a certain strength.

In one sense, the behaviour evidenced by signallers within a stable handicap signalling system is stable for the same reason that the bee dance-language is stable — the possible outcomes of the signalling scenario are ranked identically by both signallers and observers. The costs and benefits which define a handicap signalling system, define the optimal behaviour of signallers within that system. This optimal behaviour is to signal honestly.

Although signallers within a conventional signalling system might benefit (for a limited but significant period of evolutionary time) from giving a bluff signal, through reaping rewards without investing in the trait being honestly advertised by other members of the system, there is no such temptation for handicap signallers. At equilibrium, any signal other than an honest one will result in a decrease in fitness. Even a strategy which bluffs with a certain probability will simply experience a fitness decrease with that probability.

Since the making of a signal is inextricably tied to the bearing of a cost (even if this cost is in terms of increased risk, rather than increased energetic costs, and is thus not paid each time a signaller signals, but each time a signaller is probed) there is no sense in the notion that a signaller might profit from bluffing or cheating at a low frequency. There is no profit-making margin within a stable handicap signalling system.

Some authors (e.g., Johnstone, 1997; Johnstone & Grafen, 1993) have pointed out that signallers who, for some reason, are not subject to the same costs and benefits as those engaged in honest handicap signalling may benefit from bluffing. Such individuals are identical to third-party parasites of co-operative signalling scenarios. They are able to exploit an honest signalling system through some means unavailable to the co-operative signallers (e.g., through not investing in unpalatability).

Although, under this interpretation of handicap signalling, it is hardly surprising that such systems are stable, this is not the interpretation typically presented. The very term ‘handicap’ implies that such signals are somehow maladaptive, that signallers would eschew them if they could. However, the reasoning presented above casts them as eminently reasonable; as strategically sound; as the best response, given the circumstances.

In this sense, handicap signals are co-operative signals. Signallers and observers both do as well as they can do given the costs and benefits involved in the system. Signallers have been selected to signal honestly since it was through signalling honestly (rather than signalling as strongly as possible, or gaining the maximum possible observer response) that their ancestors came to do better than dishonest signallers. Thus the informative nature of the mapping between signal and signaller trait is not merely accidental. Honesty is not epiphenomenal within the system, but must feature as a substantive element of any explanation for the existence and character of any stable handicap signalling system.

7.7.4 The Generality of the Handicap Principle

The force of the results presented within this chapter is to qualify previous statements of the conditions which must be met before honest handicap signalling may be evolutionarily stable. Rather than merely requiring *gross* signalling costs to vary with signaller quality in some manner, the models presented here require that the manner in which the *net* cost of signalling varies with signaller quality be considered.

Although Zahavi often appears to consider the net costs involved in signalling when formulating his principle (e.g., “it is reasonable to expect a population in its optimal fitness to benefit from a handicap”, and “so long as the offspring . . . does not deviate to grow its handicap larger than it can afford, the handicap [may remain stable] as a marker of honest advertisement”, Zahavi, 1977a, p.604), when describing examples of natural signalling he rarely appreciates the benefits which might be accrued from signalling, and the manner in which such benefits might negate the increased costs involved in bluffing.

This ambivalence towards the potential benefits of signalling (or bluffing) led Wiley (1983) to characterise Zahavi’s (1975) claim as “signals should evolve to become a *net* handicap to signallers” (p. 176, my emphasis), whilst Adams and Mesterton-Gibbons (1995) reach the opposite conclusion stating that the difference between their model and more straightforward models of Zahavi's handicap principle resides in the fact that "the net benefit for a given advertisement may not increase monotonically with the signaller's strength" (p.406).

Furthermore, the sense of much of Zahavi's verbal argument does not seem to accord with a notion of the handicap principle couched in terms of net costs. For example, as Hurd (1995) points out, if the costs involved in signalling must be acceptable costs (i.e., they must be compensated for by consonant benefits), then in what sense are these costs a 'handicap'? Although the costs incurred by a *bluffer* might be characterised as a handicap, since these costs would not be compensated for by any observer response, this is not the sense in which Zahavi proposed the term. For Zahavi, *honest* signallers suffer a handicap. This suffering is necessary as a method of demonstrating honesty. However, once one appreciates the role played by benefits in assuaging these costs, the notion that signallers are suffering becomes incoherent.

Do dancing bees 'suffer' from the costs they incur in honestly advertising the location of nectar? Surely not. Within such co-operative signalling scenarios, any 'suffering' is more than compensated for by the benefits that such dancing will afford. Similarly, a signaller locked into an honest signalling system by the manner in which costs and benefits are mediated by some signaller trait cannot be said to *suffer* from costs given that they are more than compensated for by the benefits which they lead to. Indeed, as argued above, is there any difference between the cost-benefit conditions which ensure that handicap signalling is stable, and the cost-benefit conditions which obtain in the bee dance scenario?

The inclusion of a benefit clause in the definition of the handicap principle does not preclude the existence of handicap signalling equilibria. However, it does have implications for the proposed ubiquity of the handicap principle as it has been presented by Zahavi and others.

The conditions that (i) signals cost, and that (ii) signal cost is related to signaller quality appeared to be candidates for a very wide application. Indeed, Zahavi demonstrated the width of the application in the papers which he published, even going so far as to suggest that the handicap principle accounted for all natural signalling. However, the models constructed here demonstrate that the influence of benefits on signaller behaviour may ensure that despite signals costing, and their cost being related to signaller quality, honesty may be unstable. Similarly, some systems, despite failing to meet Zahavi's two constraints may be stable due to the influence of benefits upon signaller behaviour.

As such, the ease with which these revised conditions for the existence of evolutionarily stabile handicap signalling may be confidently predicted to hold across classes of signalling scenario is much reduced. This is compounded by the fact that as demonstrated above, theorists' predictions concerning the manner in which costs and benefits vary across signalling populations vary considerably.

From this discussion, it is clear that the models constructed within this chapter, viewed from the perspective developed within chapter 5, question the *integrity* of handicap signalling as a class of signalling distinct from co-operative signalling, and challenge its proposed *ubiquity* through highlighting the complications which arise from a consideration of the manner in which costs *and benefits* are mediated by signaller traits.

7.7.5 Extensions

The results presented here are pregnant with possibility.

Many hypotheses concerning the impact of as yet unformalised factors upon the stability of handicap signalling have been presented. Primary among these are factors concerning the role of the observer within a signalling system. Within the model presented above, an observer's fitness was considered to vary with the degree of accuracy with which she predicated the signaller trait of interest to her. It has been pointed out that where observers incur costs through their involvement in a signalling episode, the resulting complications in the calculation of observer fitness might militate against the stability of fully honest scenarios, favouring a less pronounced degree of honesty with an attendant reduction in observer costs (Dawkins & Guilford, 1991).

In implementing these observer costs, many interesting factors must be entertained. Might the costs of receiving a signal vary with the signal (e.g., louder calls increase the probability of predation by eavesdroppers)? Might the degree to which honesty may be corrupted vary with the value of the signaller trait (e.g., might systems in which females gain a more accurate assessment of low-quality suitors than high-quality suitors be stable?). Might inaccuracies in predicting signaller traits influence fitness in a manner depending on the direction of the inaccuracy? For example, might an overestimation of suitor quality reduce fitness more than an underestimation?[6] Furthermore, might the fitness consequences of over-estimation, and under-estimation vary with the actual trait possessed by the advertising suitor?

Issues such as these may be easily addressed through alterations to the fitness functions presented in section 7.6.

Maynard Smith (1985) identifies three differing selection mechanisms which might be employed by the selecting sex in a mate-choice paradigm. Results from genetic models employing these three mechanisms differ to some degree. The extent of differences resulting from alternative selection mechanisms has not been explored for game theoretic models, but there is no reason to suppose that qualitative differences between the results of models differing in this respect will not be found.

Such differing mechanisms are also easily incorporated within the evolutionary simulation model presented above.

It has been demonstrated that noise within models of handicap signalling can lead to stratification of signalling groups into discrete bands with differing mean quality. Whether this discretisation of the population leads to instability in the form of conventional signalling as discovered for the discrete model presented within this chapter is an open question which might easily be addressed by alternative versions of the models presented here.

In addition, some of the simplifying assumptions made in the construction of the models presented within this chapter may now be relaxed. The linear nature of the signalling and response strategies introduced in section 7.6 is one such simplifying assumption. The effects of non-linearities in such functions are hard to predict, but such functions are likely to better approximate the strategies employed by real signallers and receivers.

Finally, no ESS model of the handicap principle has successfully incorporated heritability of

---

[6] In the terminology of statistical hypothesis testing, might the consequences of a Type 1 error differ from those of a Type 2 error?

the advertised trait. As mentioned in chapter 5, it has been suggested (Pomiankowski, Iwasa, & Nee, 1991; Iwasa, Pomiankowski, & Nee, 1991) that biased deleterious mutation on the advertised trait may overcome the paradox of the lek through encouraging enough genetic diversity within the advertising population to maintain an observer fitness benefit in remaining choosy.

However, as was discussed in chapter 5, the model upon which this suggestion was based features some assumptions which need to be relaxed before the stability of handicap signals of heritable traits can be demonstrated convincingly. Such relaxation of assumptions might be achieved through extending the models presented here to feature heritable traits which experience mutation events of the nature proposed by Pomiankowski et al. (1991).

7.8 Conclusion

In summary, signalling equilibria were shown to exist under three conditions defined by Grafen (1990b) using the inequalities, $w_{23} < 0$, $w_{23} = 0$, and $w_{23} > 0$, and also defined for the extended version of Hurd's (1995) discrete action-response game explored here using the inequalities, $V_L < V_H$, $V_L = V_H$, and $V_L > V_H$. In concert these three classes of scenario were used to explore the effects of the benefits to signallers of their signalling behaviours, not merely the costs of such behaviours. Non-handicap signalling equilibria were shown to obtain under certain conditions. It was demonstrated that in order to show that a signalling system is stable, a relationship between signalling costs, signaller quality, and (*contra* Zahavi) signalling *benefits* must be shown to hold, *not* merely a relationship between signalling costs and signaller quality.

In addition to these analytically derived results, further exploration of Hurd's (1995) discrete action-response game was carried out utilising a simulation-based paradigm which allowed a qualitative account of the system's dynamics to be formulated. As a result of this exploration, several interesting, robust phenomena, which were opaque to the analysis carried out in section 7.2 were detailed. Amongst the phenomena described were classes of conventional signalling scenario. These stable signalling scenarios cannot be characterised as fixed points in the system's dynamics. They exist as higher dimensional attractors (e.g., limit cycles) in strategy space, and as such are not amenable to a simple ESS approach.

Further work, both analytic and simulation-based, must be undertaken before a full characterisation of the dynamics of these systems (both discrete and continuous) can be constructed, and the extent of their applicability to the evolution of natural signalling systems can be assessed.

Chapter 8

Conclusions

This thesis has sought to present, define, and motivate, a particular style of scientific modelling. This style of modelling involves the simulation of evolutionary systems as populations of individuals subject to some adaptive process. The substantive element of such *evolutionary simulation modelling*, in contrast to alternative modelling paradigms, is the ongoing, unfolding, evolutionary dynamics which evolutionary simulation models implement. Through the judicious design of such evolutionary simulation models, the character of these dynamics, including, but in no way limited to, any stable states to which the evolutionary process might lead, may allow one to discern the validity of competing theories.

8.1 A Framework for Evolutionary Simulation Modelling

A large part of this thesis has been devoted to providing a framework within which such scientific modelling may take place. In order that such a framework could be constructed it was necessary to clear the decks to a certain degree. Much of the manoeuvring which this deck clearing involved was necessary in order to distinguish evolutionary simulation modelling from, on the one hand, the general use of computers by theoretical biologists, and on the other, the recent interest in biology shown by many artificial intelligence and artificial life researchers.

Evolutionary simulation modelling certainly includes instances of the use of computers by theoretical biologists. One of these instances (Enquist & Arak, 1994) was discussed at some length within chapter 6. Similarly, evolutionary simulation modelling owes a considerable debt to those artificial intelligence and artificial life researchers inspired by biological theory. An attempt to repay some of this debt was made in chapter 4. However, from the arguments presented in chapters 2 and 3, I hope it is clear that evolutionary simulation modelling can genuinely *augment* extant theoretical tools employed in the modelling of evolutionary phenomena. Furthermore, from these same arguments, it should be clear that evolutionary simulation modelling provides a similarly genuine opportunity for artificial life researchers to practice scientific modelling, rather than mere science fiction.

The thrust of the arguments alluded to immediately above relies on some observations concerning the evolutionary process. First, evolution is a dynamic process. Furthermore, evolutionary dynamics are historical in a non-trivial manner. A system satisfying the pre-requisites for adaptation (reproduction, inheritance, variability, and competition for limited resources) is a system which will experience change of a very special kind. The change this system experiences will not be the change conceived of within teleological accounts of a static, ordered, spiritually legitimated universe. Neither will it be the change conceived of within accounts of the chaotic flux of quantum particles, or the relentless march towards cold, bland, homogeneity of increasing entropy. Indeed, the nature of the change experienced by evolutionary systems is still open for philosophical debate and empirical exploration.

What is clear is that the nature of evolutionary change is such that it gives the impression of conscious design. This feature of the evolutionary process is perhaps its most impressive attribute, and, as was argued within chapter 3, is an attribute that can only be appreciated from a historical perspective. By this I mean to make two points. The first is that an appreciation of the historical change in perspective afforded by evolutionary theory allows one to clearly assess the merits of an adaptive account of some natural phenomenon. Secondly, an appreciation of the function of an evolved system can only be gained through attending to the evolutionary history of that system. Judgements of matters such as the ability of a system to function properly may only be made within a historical context which includes the selective history of the system in question. What evolutionary simulation models offer is a scientific tool which affords theoretical primacy to the notion of undirected, yet adaptive, evolutionary change.

8.2 Examples of Evolutionary Simulation Modelling

The second part of this thesis consists of attempts to apply evolutionary simulation modelling techniques, alongside more analytical techniques, to theoretical problems current within evolutionary biology. These attempts serve both to provide prototypical instances of the application of evolutionary simulation modelling, and to provide insights into the nature of the biological problems under investigation.

The first of two biological problems addressed within this thesis was the prevalence of complex symmetrical patterns used as signals within the animal kingdom. The account of this problem presented within this thesis took the form of a cautionary tale.

An existing model (Enquist & Arak, 1994), which sought to account for the prevalence of complex symmetry as a result of the logistics of visual discrimination, was described. This model was presented by its authors in the form of an iterative computer simulation involving the application of an artificial evolutive algorithm to artificial neural networks which acted as idealised models of natural visual systems. As such the model might be regarded, with hindsight, as an evolutionary simulation model.

In chapter 6, the results of this model were questioned. It was demonstrated that the complex symmetry which results from the original model was an artefact of its construction. This demonstration provided an opportunity to raise a series of methodological issues concerning the proper design of evolutionary simulation models, and the pitfalls which may await those eager to employ evolutionary simulation models in the testing of biological hypotheses without paying due care and attention to the literature concerning the idiosyncrasies of such powerful adaptive mechanisms as genetic algorithms and artificial neural networks.

Although these tools were developed as simple analogues of their natural counterparts, they have proved to be capable of an immensely rich palette of behaviour. For example, recent classes of artificial neural network have been shown to be able to capture the profile of any dynamical system, to any degree of accuracy (Funahashi & Nakamura, 1993). Similarly, genetic algorithms have developed in many different directions, incorporating many aspects of the natural biological systems they mimic, and as a consequence encountering many of the same theoretical problems. What is the role of sex, or mutation, or sexual selection, or morphogenesis within a genetic algorithm? These are open questions within evolutionary computation partly as a result of their unresolved status within evolutionary biology.

As a result, the proper design of evolutionary simulation models will be predicated, not only on theoretical advances within evolutionary biology, but on associated advances made within evolutionary computation.

The second of the two biological problems addressed within this thesis concerned the conditions under which natural signalling might be evolutionarily stable — and received a more positive treatment. Two existing models were extended and their results, along with those of previous similar models, were compared, and contrasted. In casting the extended models as evolutionary simulation models, aspects of their behaviour which had been opaque to the prior analytical treatments were revealed, and could be qualitatively characterised.

On the basis of the findings reported in chapter 7, a reconciliation of previous results was attempted. In achieving this reconciliation, the mechanism which had purported to ensure honesty within such signalling systems (the 'handicap' principle, Zahavi, 1975, 1977a, 1977b) was demonstrated as misleading in its original conception. An alternative perspective on handicap signalling was offered, from which handicap signalling systems could be considered to belong to the less troublesome class of co-operative signalling systems (as defined in chapter 5).

8.3 Interdisciplinarity

Since evolutionary simulation models are tools, they must have tasks. Evolutionary simulation modelling involves a subject matter (e.g., the evolution of honesty, or the dynamics of economic agents within an auction). It also requires the construction of a simulation. It is thus interdisciplinary in the same sense that statistics is interdisciplinary. However, whereas statistics is taught as an integral part of the many disciplines which require its use, simulation design is, as yet, not. This ensures that an evolutionary simulation modeller is typically one of two beasts. Either she is a specialist in the field which provides the tasks for the model, or she is a specialist in the field which provides the tools for the model. This thesis will conclude with some thoughts on the implications of this situation for evolutionary simulation modelling.

As Beer rightly notes,

> "one of the excitements of undertaking work across fields is that it propels us to the edge of our own competence and makes necessary new skills and new reading" (Beer, 1991, p.4).

In applying evolutionary simulation modelling to topics with which I had erstwhile been unfamiliar, I have indeed been "propelled to the edge of my own competence". There is some satisfaction to be had in being thus propelled. Ortega y Gasset (1930) captures this satisfaction in his essay *The Barbarism of Specialisation*. For Ortega, the notion of a specialist (and a scientific specialist is a specialist *par excellence*) is tied up with the notion of an arrogant ignorance of matter lying outside a speciality. In this respect, the production of specialists is to be avoided, since they are the root of much that is wrong with society. However, in the current climate of academic compartmentalism, how is one to know whether one is *resisting* Ortega's 'barbarism' in apparently broadening one's academic perspective, or *contributing* to it, by in actual fact applying the perspective of one's own speciality to the speciality of others?

Beer (1991) presents some thoughts which speak to this issue,

> "Problems remain. How thoroughly interdisciplinary is it possible to be? Are we lightly transferring a set of terms from one practice to another, as metaphor, *facon de parler*? Are we appropriating *materials* hitherto neglected for analysis of the kind we have always used? Or are we trying to learn new *methods* and skills fast, which others have spent years acquiring? The key questions in another discipline may at first glance seem banal, since the incomer is ignorant of the resistances that have produced and shaped them. That superior glossing of other people's controversies is one hazard in interdisciplinary work; the opposite temptation is to succumb to the glamour of the horizon. Those most sceptical about their own disciplinary practices are sometimes inclined to embrace the practices of an adjacent discipline too reverentially." (Beer, 1991, p.5–6).

It appears to me that problems such as these lie at the heart of the potential for the artificial sciences to contribute to the natural sciences, i.e., to contribute to science at all.

Upon the publication of Charles Babbage's (1837) uninvited *Ninth Bridgewater Treatise*, the work was treated with some disdain. It was generally agreed to have overstepped some boundary. Much of this opprobrium seems to stem from a tacit presupposition on the part of Babbage's audience to the effect that the workings of machinery were not to offer insight into the doings of the Deity.

Like Babbage's, most contemporary evolutionary simulation models, once constructed, will typically be 'uninvited'. Linguists such as Chomsky refuse to contemplate an adaptive rationale accounting for the genesis of the Language Acquisition Device; psychologists disown those researchers attempting to provide evolutionary underpinnings for human behaviour; economists similarly deny that their theories must accommodate experimental data, let alone the results of models of economic change as susceptible to the vagaries of an adaptive historical process. The list is very long. Until formal models of these quasi-evolutionary processes can be constructed in a manner rigorous enough to quash dissenters, the "transfer of interest from the permanent to the changing" identified by Dewey (1909) as responsible for some of the major scientific revolutions will be checked. Hopefully evolutionary simulation models will be one tool with which such formal models may be constructed.

Bibliography

Adams, E. S., & Mesterton-Gibbons, M. (1995). The cost of threat displays and the stability of deceptive communication. *Journal of Theoretical Biology*, *175*, 405–421.

Andersson, M. (1980). Why are there so many threat displays?. *Journal of Theoretical Biology*, *86*, 773–781.

Andersson, M. (1982). Sexual selection, natural selection, and quality advertisement. *Biological Journal of the Linnean Society*, *17*, 375–393.

Arak, A., & Enquist, M. (1993). Hidden preferences and the evolution of signals. *Phil. Trans. R. Soc. Lond., Series* B, *340*, 207–213.

Arak, A., & Enquist, M. (1995a). Conflict, receiver bias and the evolution of signal form. *Phil. Trans. R. Soc. Lond., Series* B, *349*, 337–344.

Arak, A., & Enquist, M. (1995b). (reply to Cook). *Nature*, *374*, 313–314.

Arthur, W. B. (1993). On designing economic agents that behave like human agents. *Journal of Evolutionary Economics*, *3*, 1–22.

Ashby, W. R. (1956). *An Introduction to Cybernetics*. Wiley, New York.

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books, New York.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.

Babbage, C. (1837). *Ninth Bridgewater Treatise: A Fragment* (2nd edition). Murray.

Babbage, C. (1864). *Passages from the Life of a Philosopher*. Longman, London.

Baldwin, J. M. (1896). A new factor in evolution. *American Naturalist*, *30*, 441–451.

Baricelli, N. A. (1957). Numerical testing of evolutionary theories. *ACTA Biotheoretica*, *16*, 69–126.

Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press, New York.

Barrow, J. D., & Tipler, F. J. (1986). *The Anthropic Cosmological Principle*. Clarendon Press, Oxford.

Bartel, D. P., & Szostak, J. W. (1993). Isolation of new ribozymes from a large pool of random sequences. *Science*, *261*, 1411–1418.

Bedau, M. (1996). The nature of life. In Boden, M. A. (Ed.), *The Philosophy of Artificial Life*, pp. 332–357. Oxford University Press, Oxford.

Beer, G. (1983). *Darwinian Plots: Evolutionary Narrative in Darwin, George Eliot and Nineteenth-Century Fiction*. Routledge & Kegan Paul, London.

Beer, G. (1991). *Forging the missing link — Interdisciplinary Stories*. Cambridge University Press, Cambridge.

Beer, R. D. (1990). *Intelligence as Adaptive Behavior: An experiment in computational neuroethology*. Academic Press, Boston.

Bekoff, M., & Allen, C. (1992). Intentional icons: Towards an evolutionary cognitive ethology. *Ethology*, *91*, 1–16.

Best, M. L. (1997). Models for interacting populations of memes: Competition and niche behavior. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97), Brighton, UK, 28–31 July*, Complex Adaptive Systems Series, pp. 154–162 Cambridge, MA. MIT Press.

Bickerton, D. (1990). *Language and Species*. University of Chicago Press, Chicago.

Binmore, K. (1987). Modelling rational players, part I. *Economics and Philosophy*, *3*, 179–214.

Binmore, K. (1988). Modelling rational players, part II. *Economics and Philosophy*, *4*, 9–55.

Binmore, K. (1991). *Fun and Games*. D. C. Heath, Lexington, MA.

Binmore, K. (1994). *Playing Fair: Game Theory and the Social Contract*. MIT Press, Cambridge, MA.

Boden, M. (1996a). Autonomy and artificiality. In Boden, M. A. (Ed.), *The Philosophy of Artificial Life*, pp. 95–108. Oxford University Press, Oxford.

Boden, M. A. (Ed.). (1996b). *The Philosophy of Artificial Life*. Oxford University Press.

Bonabeau, E. W., & Theraulaz, G. (1996). Why do we need artificial life?. In Langton, C. G. (Ed.), *Artificial Life: An Overview*, pp. 303–325. MIT Press: A Bradford Book.

Bowler, P. J. (1989). *The Invention of Progress: The Victorians and the Past*. Basil Blackwell, Oxford.

Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. The University of Chicago Press, Chicago.

Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA.

Briscoe, T. (1997). Language acquisition: The bioprogram hypothesis and the Baldwin effect. MS, Computer Laboratory, University of Cambridge.

Brock, W. H. (1966). The selection of the authors of the Bridgewater Treatises. *Notes and Records of the Royal Society of London*, *21*, 162–179.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

Bullock, J. (1994). Letter. *Trends in Ecology and Evolution*, *9*(8), 299.

Bullock, S. (1995a). Co-evolutionary design: Implications for evolutionary robotics. CSRP 384, University of Sussex.

Bullock, S. (1995b). Dynamic fitness landscapes. In de Bourcier, P., Lemmen, R., & Thompson, A. (Eds.), *The Seventh White House Papers: Graduate research in the Cognitive and Computing Sciences at Sussex*. University of Sussex.

Bullock, S. (1997). An exploration of signalling behaviour by both analytic and simulation means for both discrete and continuous models. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97), Brighton, UK, 28–31 July*, pp. 454–463 Cambridge, MA. MIT Press / Bradford Books.

Bullock, S., & Cliff, D. (1997). The role of 'hidden preferences' in the artificial co-evolution of symmetrical signals. *Proc. R. Soc. Lond., Series* B, *264*, 505–511.

Burk, T. (1988). Acoustic signals, arms races and the costs of honest signalling. *Florida Entomologist*, *71*(4), 400–409.

Butterworth, G. E., Rutkowska, J. C., & Scaife, M. (Eds.). (1985). *Evolution and Developmental Psychology*. Harvester Press, Brighton.

Caryl, W. (1979). Communication by agonistic displays: What can games theory contribute to ethology?. *Behavior*, *68*, 136–169.

Caryl, W. (1981). Escalated fighting and the war of nerves: Games theory and animal combat. In Bateson, P. P. G., & Klopfer, P. H. (Eds.), *Perspectives in Ethology*, Vol. 4, pp. 199–224. Plenum Press, New York.

Caryl, W. (1982). Animal signals: A rely to Hinde. *Animal Behaviour*, *30*, 240–244.

Churchland, P. M. (1984). *Matter and Consciousness*. MIT Press, Cambridge, MA.

Clark, A. C. (1989). *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. MIT Press, Cambridge, MA.

Clark, A. (1990). Connectionism, competence and explanation. In Boden, M. A. (Ed.), *The Philosophy of Artificial Intelligence*, pp. 281–308. Oxford University Press.

Cliff, D. (1991). Computational neuroethology: A provisional manifesto. In Meyer, J.-A., & Wilson, S. W. (Eds.), *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior (SAB'90)*, pp. 29–39 Cambridge, MA. MIT Press Bradford Books.

Cliff, D., & Bullock, S. (1993). Adding "Foveal Vision" to Wilson's Animat. *Adaptive Behaviour*, *2*(1), 49–72.

Cliff, D., Harvey, I., & Husbands, P. (1993). Explorations in evolutionary robotics. *Adaptive Behaviour*, *2*(1), 71–108.

Cliff, D., & Miller, G. F. (1995). Tracking the Red Queen: Measurements of adaptive progress in co-evolutionary simulations. In Morán, F., Moreno, A., Morelo, J. J., & Chacón, P. (Eds.), *Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life*. Springer.

Cliff, D., & Miller, G. F. (1996). Co-evolution of pursuit and evasion II: Simulation and results. In Maes, P., Mataric, M., Meyer, J.-A., Pollack, J., & Wilson, S. W. (Eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on the Simulation of Adaptive Behaviour*, pp. 506–515. MIT Press/Bradford Books.

Clutton-Brock, T. H., & Albon, S. D. (1979). The roaring of red deer and the evolution of honest advertisement. *Behaviour*, *69*, 145–170.

Coker, P., & Winter, C. (1997). N-sex reproduction in dynamic environments. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97), Brighton, UK, 28–31 July*, pp. 163–169 Cambridge, MA. MIT Press / Bradford Books, Cambridge, MA.

Colasanti, R., & Hunt, R. (1997). Real botany with artificial plants: A dynamic self assembling plant model for individuals and populations. In Husbands, P., & Harvey, I. (Eds.), *Fourth European Conference on Artificial Life*, pp. 266–273. MIT Press.

Cook, N. D. (1995). Artefact or network evolution. *Nature, 374,* 313.

Costall, A. (1985). Specious origins? Darwinism and developmental theory. In Butterworth, G. E., Rutkowska, J. C., & Scaife, M. (Eds.), *Evolution and Developmental Psychology,* pp. 30–41. Harvester Press, Brighton.

Darwin, C. R. (1859). *On the Origin of Species*. John Murray, London.

Darwin, C. R. (1871). *The Descent of Man, and Selection in Relation to Sex*. John Murray, London.

Davies, N. B., Bourke, A. F. G., & de L. Brooke, M. (1989). Cuckoos and parasitic ants: Interspecific brood parasitism as an evolutionary arms race. *Trends in Ecology and Evolution*, *4*, 274–278.

Davis, J. W. F., & O'Donald, P. (1976). Sexual selection for a handicap: A critical analysis of Zahavi's model. *Journal of Theoretical Biology*, *57*, 345–354.

Dawid, H. (1996). *Adaptive Learning by Genetic Algorithms: Analytical Results and Applications to Economic Models*. No. 441 in Lecture Notes in Economics and Mathematical Systems. Springer.

Dawkins, M. S. (1995). *Unravelling Animal Behaviour*. Longman, Harlow.

Dawkins, M. S., & Guilford, T. (1991). The corruption of honest signalling. *Animal Behaviour*, *41*, 865–873.

Dawkins, M. S., & Guilford, T. (1995). An exaggerated preference for simple neural network models of signal evolution?. *Proc. R. Soc. Lond., Series* B, *261*, 357–360.

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.

Dawkins, R. (1986). *The Blind Watchmaker*. Longman, Harlow.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation. In Krebs, J. R., & Davies, N. B. (Eds.), *Behavioural Ecology: An Evolutionary Approach*, pp. 282–309. Blackwell, Oxford.

Dawkins, R., & Krebs, J. R. (1979). Arms races between and within species. *Proc. R. Soc. Lond., Series* B, *205*, 489–511.

de Bourcier, P., & Wheeler, M. (1994). Signalling and territorial aggression: An investigation by means of Synthetic Behavioural Ecology. In Cliff, D., Husbands, P., Meyer, J.-A., & Wilson, S. W. (Eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on the Simulation of Adaptive Behaviour*, pp. 436–472. MIT Press/Bradford Books.

de Bourcier, P., & Wheeler, M. (1995). Aggressive signalling meets adaptive receiving: Further experiments in Synthetic Behavioural Ecology. In Morán, F., Moreno, A., Morelo, J. J., & Chacón, P. (Eds.), *Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life* Berlin and Heidelberg. Springer-Verlag.

DeAngelis, D., & Gross, L. (Eds.). (1992). *Individual Based Models and Approaches in Ecology: Populations, Communities, and Ecosystems*. Chapman and Hill, New York, NY.

Dewey, J. (1909). The influence of Darwinism on philosophy. In Gardner, M. (Ed.), *The Sacred Beetle and other great essays in science*, pp. 18–31. Oxford University Press, Oxford. This collection published in 1984.

di Paolo, E. A. (1996a). An investigation into the evolution of communicative behaviors. Cognitive science research paper 445, School of Cognitive and Computing Sciences, University of Sussex.

di Paolo, E. A. (1996b). Some false starts in the construction of a research methodology for artificial life. In Noble, J., & Parsowith, S. (Eds.), *The 9th White House papers: Graduate Research in the Cognitive & Computing Sciences at Sussex*. University of Sussex, School of Cognitive and Computing Sciences, Brighton, U.K.

Dosi, G., & Nelson, R. R. (1994). An introduction to evolutionary theories in economics. *Journal of Evolutionary Economics*, *4*, 153–172.

Dreyfus, H. L. (1979). *What computers can't do: The limits of artificial intelligence*. Harper, New York. Rev. ed.: First published as: What computers can't do: A critique of artificial reason, 1972.

Eco, U. (Ed.). (1976). *A Theory of Semiotics*. Indiana University Press, Bloomington.

Elton, M. (1994). What are you talking about?. *AISB Quarterly*, *87*, 47–54. Special Theme: AI and Artificial Life, ed. D. Cliff.

Endler, J. A. (1991). Interactions between predators and prey. In Krebs, J. R., & Davies, N. B. (Eds.), *Behavioural Ecology: An Evolutionary Approach* (3rd edition)., pp. 169–201. Blackwell, Oxford.

Endler, J. A. (1993). Some general comments on the evolution and design of animal communication systems. *Phil. Trans. R. Soc. Lond., Series* B, *340*, 215–225.

Enquist, M. (1985). Communication during aggressive interactions with particular reference to variation in choice of behaviour. *Animal Behaviour*, *33*, 1152–1161.

Enquist, M., & Arak, A. (1993). Selection of exaggerated male traits by female aesthetic senses. *Nature*, *361*, 446–448.

Enquist, M., & Arak, A. (1994). Symmetry, beauty and evolution. *Nature*, *372*, 169–172.

Epstein, J., & Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press, Brookings, MA.

Eshel, I. (1978). On the handicap principle — a critical defence. *Journal of Theoretical Biology*, *70*, 245–250.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Fodor, J. A. (1975). *The Language of Thought*. Harvester Press.

Fontana, W., Wagner, G., & Buss, L. W. (1994). Beyond digital naturalism. *Artificial Life*, *1*(1), 211–227.

Foster, J. (1989). *Evolutionary Macroeconomics*. Unwin-Hyman, Boston.

Franceshini, N., Pichon, J. M., & Blanes, C. (1992). From insect vision to robot vision. *Phil. Trans. R. Soc. Lond., Series* B, *337*, 283–294.

Fraser, A. S. (1957a). Simulation of genetic systems by automatic digital computers: I. Introduction. *Australian Journal of Biological Science*, *10*, 484–491.

Fraser, A. S. (1957b). Simulation of genetic systems by automatic digital computers: II. Effects of linkage on rates of advance under selection. *Australian Journal of Biological Science*, *10*, 492–499.

Funahashi, K., & Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, *6*, 801–806.

Gale, J. S., & Eaves, L. J. (1975). Logic of animal conflict. *Nature*, *254*, 463–464.

Gode, D. K., & Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, *101*(1), 119–137.

Godfray, H. C. J. (1991). Signalling of need by offspring to their parents. *Nature*, *352*, 328–330.

Godfray, H. C. J. (1995). Evolutionary theory of parent-offspring conflict. *Nature*, *376*, 133–138.

Godfrey-Smith, P. (1996). Spencer and Dewey on life and mind. In Boden, M. A. (Ed.), *The Philosophy of Artificial Life*, pp. 314–331. Oxford University Press, Oxford.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.

Gopnik, M. (1994). From families to phenotypes: Research into the genetic basis of specific language impairment. In Watkins, R., & Rice, M. (Eds.), *New directions in specific language impairment*. Baltimore, MD.

Gould, S. J. (1977). Eternal metaphors of palaeontology. In Hallam, A. (Ed.), *Patterns of Evolution*, pp. 1–26. Elsevier, Amsterdam.

Gould, S. J. (1989). *Wonderful Life: The Burgess Shale and then Nature of History*. W. W. Norton, New York.

Gould, S. J. (1996). *Full House: The Spread of Excellence from Plato to Darwin*. Harmony Books.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marcos and the Panglossian program: A critique of the adaptationist programme. *Proc. R. Soc. Lond., Series* B, *205*, 281–288.

Grafen, A. (1990a). Sexual selection unhandicapped by the Fisher process. *Journal of Theoretical Biology*, *144*, 473–516.

Grafen, A. (1990b). Biological signals as handicaps. *Journal of Theoretical Biology*, *144*, 517–546.

Grafen, A., & Johnstone, R. A. (1993). Why we need ESS signalling theory. *Phil. Trans. R. Soc. Lond., Series* B, *340*, 245–250.

Grand, S., Cliff, D., & Malhotra, A. (1997). Creatures: Artificial life autonomous software agents. In *Proceedings of the First International Conference on Autonomous Software Agents (Agents97)*. In press.

Gray, F. (1953). Pulse code communication. U.S. Patent 2 632 058.

Gribbin, J. (1997). Darwin in space. *Prospect*, *May*.

Griffin, D. R. (1976). *The Question of Animal Awareness: Evolutionary Continuity of Mental Experience*. Rockefeller University Press.

Griffin, D. R. (1984). *Animal Thinking*. Harvard University Press, Cambridge, MA.

Griffin, D. R. (1992). *Animal Minds*. University of Chicago Press, Chicago.

Guilford, T., & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour*, *42*, 1–14.

Guilford, T., & Dawkins, M. S. (1993). Receiver psychology and the design of animal signals. *Trends in Neurosciences*, *16*(2), 430–436.

Haeckel, E. H. (1874). *The Evolution of Man*. Watts, London. Translated from the German 5th edition, 1905.

Hamilton, W. D. (1964). The genetical evolution of social behaviour, I and II. *Journal of Theoretical Biology*, *7*, 1–16; 17–32.

Hamilton, W. D., & Zuk, M. (1982). Heritable true fitness and bright birds: A role for parasites?. *Science*, *218*, 384–387.

Harvey, I. (1993). The puzzle of the persistent question marks: A case study in genetic drift. In Forrest, S. (Ed.), *Genetic Algorithms: Proceedings of the Fifth International Conference (GA93)*, pp. 15–22 San Mateo, CA. Morgan Kaufman.

Harvey, I., Husbands, P., & Cliff, D. (1993). Issues in evolutionary robotics. In Meyer, J.-A., Roitblat, H., & Wilson, S. W. (Eds.), *From Animals to Animats 2: Proceedings of the Second International Conference on Simulation of Adaptive Behaviour*, pp. 364–373 Cambridge, MA. MIT Press/Bradford Books.

Harvey, I., Husbands, P., & Cliff, D. (1994). Seeing the light: Artificial evolution, real vision. In Cliff, D., Husbands, P., Meyer, J.-A., & Wilson, S. W. (Eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behaviour*, pp. 392–401 Boston, MA. MIT Press/Bradford Books.

Hasson, O. (1991). Pursuit deterrent signals: The communication between prey and predator. *Trends in Ecology and Evolution*, *6*, 325–329.

Hasson, O. (1994). Cheating signals. *Journal of Theoretical Biology*, *167*, 223–238.

Hauser, M. D. (1996). *The Evolution of Communication*. MIT Press / Bradford Books, Cambridge, MA.

Heidegger, M. (1926). *Being and Time*. Blackwell, Oxford. Translated by J. Macquarrie and E. Robinson, 1962.

Hendriks-Jansen, H. (1996). *Catching Ourselves In The Act: Situated Activity, Interactive Emergence, Evolution, and Human Thought*. MIT Press, Cambridge, MA.

Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve : Intelligence and Class Structure in American Life*. Free Press.

Hillis, D. (1990). Co-evolving parasites improve simulated evolution as an optimisation procedure. *Physica D*, *42*, 228–234.

Hinde, R. A. (1981). Animal signals: Ethological and games-theory approaches are not incompatible. *Animal Behaviour*, *29*, 535–542.

Hinton, G. E., & Nowlan, S. J. (1987). How learning can guide evolution. *Complex Systems*, *1*, 495–502.

Hodgson, G. M. (1991). Economic evolution: Intervention contra Pangloss. *Journal of Economic Issues*, *25*(2), 519–533.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbour. Reprinted by MIT Press, 1992.

Hurd, P. L. (1995). Communication in discrete action-response games. *Journal of Theoretical Biology*, *174*, 217–222.

Hurd, P. L. (1997). Is signalling of fighting ability costlier for weaker individuals?. *Journal of Theoretical Biology*, *184*, 83–88.

Hurd, P. L., Wachtmeister, C.-A., & Enquist, M. (1995). Darwin's principle of antithesis revisited: A role for perceptual biases in the evolution of intraspecific signals. *Proc. R. Soc. Lond., Series* B, *259*, 201–205.

Husbands, P. (1992). Genetic algorithms in adaptation and optimisation. In Kronkski, L. (Ed.), *Advances in Parallel Algorithms*. Blackwell Scientific.

Husbands, P. (1993). An ecosystems model for integrated production planning. *International Journal of Computer Integrated Manufacturing*, *6*(1 & 2), 74–86.

Huston, M., DeAngelis, D., & Post, W. (1988). New computer models unify ecological theory. *Bioscience*, *38*(10), 682–691.

Hutzler, G., Gortais, B., & Drogul, A. (1997). The garden of chances: An integrated approach to abstract painting and reactive DAI. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97), Brighton, UK, 28–31 July*, Complex Adaptive Systems Series, pp. 566–573 Cambridge, MA. MIT Press.

Iwasa, Y., Pomiankowski, A., & Nee, S. (1991). The evolution of costly mate preferences II: The "handicap" principle. *Evolution*, *45*(6), 1431–1442.

Jackendoff, R. (1993). *Patterns in the Mind: Language and Human Nature*. Harvester Press, Brookings, MA.

Janzen, D. H. (1980). When is it co-evolution?. *Evolution*, *34*(3), 611–612.

Johnstone, R. A. (1994a). Female preference for symmetrical males as a by-product of selection for mate recognition. *Nature*, *372*, 172–175.

Johnstone, R. A. (1994b). Honest signalling, perceptual error and the evolution of 'all-or-nothing' displays. *Proc. R. Soc. Lond., Series* B, *256*, 169–175.

Johnstone, R. A. (1995). (reply to Cook). *Nature*, *374*, 313.

Johnstone, R. A. (1997). The evolution of animal signals. In Krebs, J. R., & Davies, N. B. (Eds.), *Behavioural Ecology: An Evolutionary Approach* (4th edition)., pp. 155–178. Blackwell, Oxford.

Johnstone, R. A., & Grafen, A. (1992a). The continuous Sir Philip Sidney Game: A simple model of biological signalling. *Journal of Theoretical Biology*, *156*, 215–234.

Johnstone, R. A., & Grafen, A. (1992b). Error-prone signalling. *Proc. R. Soc. Lond., Series* B, *248*, 229–233.

Johnstone, R. A., & Grafen, A. (1993). Dishonesty and the handicap principle. *Animal Behaviour*, *46*, 759–764.

Jongeling, T. B. (1996). Self-organization and competition in evolution: A conceptual problem in the use of fitness landscapes. *Journal of Theoretical Biology*, *178*, 369–373.

Judson, O. P. (1994a). (Reply to J. Bullock). *Trends in Ecology and Evolution*, *9*(8), 299.

Judson, O. P. (1994b). The rise of the individual-based model in ecology. *Trends in Ecology and Evolution*, *9*(1), 9–14.

Kauffman, S. A. (1989). Cambrian explosion and permian quiescence: Implications of a rugged fitness landscape. *Evolutionary Ecology*, *3*, 274–281.

Kauffman, S. A. (1993). *The Origins of Order: Self-Organisation and Selection in Evolution*. Oxford University Press, New York.

Kirby, S., & Hurford, J. (1997). Learning, culture and evolution in the origin of linguistic constraints. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97), Brighton, UK, 28–31 July*, pp. 493–502 Cambridge, MA. MIT Press / Bradford Books, Cambridge, MA.

Kirkpatrick, M. (1986). The handicap mechanism of sexual selection does not work. *American Naturalist*, *127*, 222–240.

Kirkpatrick, M., & Ryan, M. J. (1991). The evolution of mating preferences and the paradox of the lek. *Nature*, *350*, 33–38.

Kitano, H., Hamahasni, S., Kitazawa, J., Takao, K., & Imai, S. (1997). The virtual biology laboratories: A new approach of computational biology. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97), Brighton, UK, 28–31 July*, pp. 274–283 Cambridge, MA. MIT Press.

Koza, J. R. (Ed.). (1993). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA.

Krakauer, D. C., & Johnstone, R. A. (1995). The evolution of exploitation and honesty in animal communication: A model using artificial neural networks. *Phil. Trans. R. Soc. Lond., Series* B, *348*, 355–361.

Krakauer, D. C., & Pagel, M. (1995). Spatial structure and the evolution of honesty cost-free signalling. *Proc. R. Soc. Lond., Series* B, *260*, 365–372.

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind reading and manipulation. In Krebs, J. R., & Davies, N. B. (Eds.), *Behavioural Ecology: An Evolutionary Approach* (2nd edition)., pp. 380–402. Blackwell, Oxford.

Lamarck, J. B. P. A. d. M. d. (1809). *Zoological Philosophy: An exposition with regard to the natural history of animals*. University of Chicago Press, Chicago. Translated from the French "Philosophie Zoologique" by Hugh Elliot, 1984.

Langton, C. G. (1991). Life at the edge of chaos. In Langton, C. G., Taylor, C., Farmer, J. D., & Rasmussen, S. (Eds.), *Artificial Life II — SFI Studies in the Sciences of Complexity*, Vol. X, pp. 41–91 Redwood City, California. Addison-Wesley.

Langton, C. G. (1996). Artificial life. In Boden, M. A. (Ed.), *The Philosophy of Artificial Life*, pp. 39–94. Oxford University Press, Oxford.

Lewontin, R. C., Rose, S., & Kamin, L. (1984). *Not in Our Genes: Biology, Ideology, and Human Nature*. Random House, New York.

Lindgren, K., & Nordahl, M. G. (1994). Evolutionary dynamics of spatial games. *Physica* D, *75*, 292–309.

MacLennan, B. (1991). Synthetic Ethology: An approach to the study of communication. In Langton, C. G., Taylor, C., Farmer, J. D., & Rasmussen, S. (Eds.), *Artificial Life II — SFI Studies in the Sciences of Complexity*, Vol. X, pp. 631–658 Redwood City, California. Addison-Wesley.

Malthus, T. (1798). *An Essay on the Principle of Population.* Patricia James (Ed.), 1989, Cambridge University Press, Cambridge.

Markl, H. (1985). Manipulation, modulation, information, cognition: Some of the riddles of communication. In Hölldobler, B., & Lindauer, M. (Eds.), *Experimental Behavioral Ecology and Sociobiology*, pp. 163–194. Sindauer, Sunderland, MA.

Marr, D. (1977). Artificial intelligence — a personal view. In Boden, M. A. (Ed.), *The Philosophy of Artificial Intelligence*, pp. 133–147. Oxford University Press, Oxford. Collection published in 1990.

Matthews, G. B. (1996). Aristotle on life. In Boden, M. A. (Ed.), *The Philosophy of Artificial Life*, pp. 303–313. Oxford University Press, Oxford.

Mayley, G. (1996). Landscapes, learning costs and genetic assimilation. *Evolutionary Computation*, *4*(3), 213–234. Special Issue on Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect, edited by P. Turney, D. Whitley & R. Anderson.

Mayley, G. (1997). Guiding or hiding: Eplorations into the effects of learning on the rate of evolution. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97), Brighton, UK, 28–31 July*, pp. 135–144 Cambridge, MA. MIT Press / Bradford Books, Cambridge, MA.

Maynard Smith, J. (1970). Time in the evolutionary process. *Studies in Genetics*, *23*, 266–272.

Maynard Smith, J. (1974a). *Models in Ecology*. Cambridge University Press, Cambridge.

Maynard Smith, J. (1974b). The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology*, *47*, 209–221.

Maynard Smith, J. (1976). Sexual selection and the handicap principle. *Journal of Theoretical Biology*, *57*, 239–242.

Maynard Smith, J. (1978). The handicap principle — a comment. *Journal of Theoretical Biology*, *70*, 251–252.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press, Cambridge.

Maynard Smith, J. (1985). Mini review: Sexual selection, handicaps and true fitness. *Journal of Theoretical Biology*, *115*, 1–8.

Maynard Smith, J. (1991). Honest signalling: The Philip Sidney game. *Animal Behaviour*, *42*, 1034–1035.

Maynard Smith, J. (1994). Must reliable signals always be costly?. *Animal Behaviour*, *47*, 1115–1120.

Maynard Smith, J. (1996). The games lizards play. *Nature*, *380*, 198–199.

Maynard Smith, J., & Harper, D. G. C. (1995). Animal signals: Models and terminology. *Journal of Theoretical Biology*, *177*, 305–311.

Maynard Smith, J., & Parker, G. A. (1976). The logic of asymmetric contests. *Animal Berhaviour*, *24*, 159–175.

Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, *246*, 15–18.

Maynard Smith, J., & Riechert, S. (1984). A conflicting-tendency model of spider agonistic behaviour: Hybrid-pure population line comparisons. *Animal Behaviour*, *32*, 564–578.

Maynard Smith, J., & Szathmáry, E. (1995). *The Major Transitions in Evolution*. Freeman, Oxford.

Maynard Smith, J., & Szathmáry, E. (1996). On the likelihood of habitable worlds. *Nature*, *384*, 107.

McCauley, E., Wilson, W. G., & de Roos, A. M. (1993). Dynamics of age-structured and spatially structured predator-prey interactions: Individual-based models and population-level formulations. *American Naturalist*, *142*, 412–442.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group (Eds.). (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 2. MIT Press/Bradford Books, Cambridge, MA.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.

Metz, J. A. J., & de Roos, A. M. (1992). The role of physiologically structured population models within a general individual-based modelling perspective. In DeAngelis, D., & Gross, L. (Eds.), *Individual Based Models and Approaches in Ecology: Populations, Communities, and Ecosystems*, pp. 88–111. Chapman and Hill, New York, NY.

Miller, G. F. (1994). Exploiting mate choice in evolutionary computation: Sexual selection as a process of search, optimization, and diversification. In Fogarty, T. C. (Ed.), *Evolutionary Computing: Proceedings of the 1994 Artificial Intelligence and Simulation of Behaviour (AISB) Society Workshop*, pp. 65–79. Springer-Verlag.

Miller, G. F. (1995). Artificial life as theoretical biology: How to do real science with computer simulation. Tech. rep., University of Sussex Cognitive and Computing Sciences Technical Report CSRP 378.

Miller, G. F., & Cliff, D. (1994). Protean behavior in dynamic games: Arguments for the coevolution of pursuit-evasion tactics. In Cliff, D., Husbands, P., Meyer, J.-A., & Wilson, S. W. (Eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on the Simulation of Adaptive Behaviour*, pp. 411–420. MIT Press/Bradford Books.

Miller, G. F., & Todd, P. M. (1990). Exploring adaptive agency I: Theory and methods for simulating the evolution of learning. In Touretsky, D. S., Elman, J. L., Sejnowski, T. J., & Hinton, G. E. (Eds.), *Proceedings of the 1990 Connectionist Models Summer School*, pp. 65–80 San Mateo, CA. Morgan Kaufmann.

Miller, J. (1997). Start the week. BBC Radio 4. Discussion of the television documentary 'Nazis: A warning from history".

Millikan, R. G. (1984). *Language, Thought and Other Biological Categories*. Cambridge University Press.

Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science*, *56*(2), 288–302.

Millikan, R. G. (1993a). *White Queen Psychology and Other Essays for Alice.* MIT Press.

Millikan, R. G. (1993b). *On mentalese orthography, Part 1*, pp. 103–122. MIT Press.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. MIT Press Bradford Books, Cambridge, MA.

Morton, P. (1984). *The Vital Science: Biology and Literary Imagination, 1860–1900*. Alen and Unwin, London.

Murdoch, W. W., McCauley, E., Nisbet, R. M., Gurney, W. S. C., & de Roos, A. M. (1992). Individual-based models: Combining testability and generality. In DeAngelis, D., & Gross, L. (Eds.), *Individual Based Models and Approaches in Ecology: Populations, Communities, and Ecosystems*, pp. 88–111. Chapman and Hill, New York, NY.

Nelson, R. R., & Winter, S. G. (1974). Neoclassical vs. evolutionary theories of economic growth: Critique and prospectus. *Economic Journal*, *84*(4), 886–905.

Nelson, R. R., & Winter, S. G. (1982). *An Evolutionary Theory of Economic Change.* Harvard University Press, Cambridge, MA.

Newell, A., & Simon, H. A. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the Association for Computing Machinery*, *19*.

Nitecki, M. H. (1988). *Evolutionary Progress.* University of Chicago Press, Chicago.

Noble, J. (1997). The scientific status of artificial life. Poster presented at the Fourth European Conference on Artificial Life (ECAL'97), Brighton, UK, 28–31 July.

Nowak, M. A., & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature*, *359*, 826–829.

Nur, N., & Hasson, O. (1984). Phenotypic plasticity and the handicap principle. *Journal of Theoretical Biology*, *110*, 275–297.

Ortega y Gasset, J. (1930). *The Barbarism of "Specialisation"*, chap. XII. Allen & Unwin, London. This translation of *La rebelion de las masas* published in 1961.

Osorio, D. (1996). Symmetry detection by categorisation of spatial phase, a model. *Proc. R. Soc. Lond., Series* B, *263*, 105–110.

Paley, W. (1802). *Natural Theology, or Evidence of the Existence and Attributes of the Deity Collected from the Appearances of Nature.* Faulder, London.

Paredis, J. (1995). Co-evolutionary computation. *Artificial Life*, *2*(4), 355–375.

Pattee, H. H. (1996). Simulations, realizations, and theories of life. In Boden, M. A. (Ed.), *The Philosophy of Artificial Life*, pp. 379–393. Oxford University Press, Oxford.

Peacocke, C. (1986). Explanation in computational psychology: Language, perception and level 1.5. *Mind and Language*, *1*(2), 101–123.

Petrie, M. (1994). Improved growth and survival of offspring of peacocks with more elaborate trains. *Nature*, *371*, 598–599.

Petrie, M., Halliday, T., & Saunders, C. (1991). Peahens prefer peacocks with elaborate trains. *Animal Behaviour*, *41*, 323–331.

Piaget, J. (1972). *The Principles of Genetic Epistemology*. Routledge & Kegan Paul, London.

Pinker, S. (1994). *The Language Instinct: The New Science of Language and Mind*. Penguin Press, London.

Pollack, J., Blair, A., & Land, M. (1996). Coevolution of a backgammon player. In Langton, C. G., & Shimohara, T. (Eds.), *Artificial Life V — Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems* Cambridge, MA. MIT Press.

Pomiankowski, A. (1987). Sexual selection: The handicap principle does work sometimes. *Proc. R. Soc. Lond., Series* B, *231*, 123–145.

Pomiankowski, A., Iwasa, Y., & Nee, S. (1991). The evolution of costly mate preferences I: Fisher and biased mutation. *Evolution*, *45*(6), 1422–1430.

Pomiankowski, A., & Møller, A. P. (1995). A resolution of the lek paradox. *Proc. R. Soc. Lond., Series* B, *260*, 21–29.

Raup, D. M. (1988). Testing the fossil record for evolutionary progress. In Nitecki, M. H. (Ed.), *Evolutionary Progress*, pp. 293–317. University of Chicago Press, Chicago.

Ray, T. (1992). Evolution, ecology and optimization of digital organisms. Working paper 92-08-042, Santa Fe Institute.

Ray, T. (1994). An evolutionary approach to synthetic biology. *Artificial Life*, *1*(1), 185–215.

Rechenberg, I. (Ed.). (1973). *Evolutionstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution*. Frommann-Holzboorg, Stuttgart.

Reynolds, C. W. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, *21*, 25–34.

Reynolds, C. W. (1994). Competition, coevolution and the game of tag. In Brooks, R., & Maes, P. (Eds.), *Artificial Life IV — Proceedings of the Workshop on the Synthesis and Simulation of Living Systems*, pp. 59–69 Cambridge, MA. MIT Press.

Ristau, C. A. (Ed.). (1991). *Cognitive Ethology: The Minds of Other Animals — Essays in honor of Donald R. Griffin*. Lawrence Erlbaum, Hillsdale, N.J.

Robbins, P. (1994). The effect of parasitism on the evolution of a communication protocol in an artificial life simulation. In Cliff, D., Husbands, P., Meyer, J.-A., & Wilson, S. W. (Eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behaviour*. MIT Press/Bradford Books;.

Robson, J. M. (1990). The fiat and the finger of God: The Bridgewater Treatises. In Helmstadter, R. J., & Lightman, B. (Eds.), *Victorian Crisis in Faith: Essays on continuity and change in 19th century religious belief*. Macmillan, Basingstoke.

Romanes, G. J. (1883). *Animal Intelligence*. Appleton.

Romanes, G. J. (1888). *Mental Evolution in Man*. Kegan Paul, Trench and Co., London.

Rosin, C. D. (1997). Coevolutionary search among adversaries. Unpublished phd thesis, University of California, San Diego, Department of Computer Science.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (Eds.). (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1. MIT Press/Bradford Books, Cambridge, MA.

Ruse, M. (1985). *Sociobiology: Sense or nonsense?* (2nd edition). Reidel, Dordrecht.

Ruse, M. (1988). Molecules to man: Evolutionary biology and thoughts of progress. In Nitecki, M. H. (Ed.), *Evolutionary Progress*, pp. 97–126. University of Chicago Press, Chicago.

Rutkowska, J. (1994). Emergent functionality in human infants. In Cliff, D., Husbands, P., Meyer, J.-A., & Wilson, S. W. (Eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behaviour*. MIT Press/Bradford Books.

Ryan, M. J. (1990). Sexual selection, sensory systems, and sensory exploitation. *Oxford Survey of Evolutionary Biology*, *7*, 157–195.

Ryan, M. J. (1997). Sexual selection and mate choice. In Krebs, J. R., & Davies, N. B. (Eds.), *Behavioural Ecology: An Evolutionary Approach* (4th edition)., pp. 179–202. Blackwell, Oxford.

Ryan, M. J., & Rand, A. S. (1993). Sexual selection and signal evolution: The ghost of biases past. *Phil. Trans. R. Soc. Lond., Series* B, *340*, 187–195.

Sanchez, E., & Tomassini, M. (Eds.). (1996). *Towards Evolutionary Hardware: The Evolutionary Engineering Approach*. Springer-Verlag.

Sasaki, T., & Tokoro, M. (1997). Adaptation towards changing environments: Why Darwinian in nature?. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97), Brighton, UK, 28–31 July*, pp. 145–153 Cambridge, MA. MIT Press / Bradford Books, Cambridge, MA.

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, *3*, 417–457.

Sebeok, T. A. (Ed.). (1977). *How Animals Communicate*. Indiana University Press, Bloomington.

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.

Sims, K. (1995). Evolving 3D morphology and behavior by competition. *Artificial Life*, *1*, 353–372.

Sinervo, B., & Lively, C. M. (1996). The rock-paper-scissors game and the evolution of alternative male strategies. *Nature*, *380*, 240.

Smith, W. J. (1977). *The Behavior of Communicating*. Harvard University Press, Cambridge, MA.

Smithers, T. (1992). Taking eliminative materialism seriously: A methodology for autonomous systems research. In Varela, F. J., & Bourgine, P. (Eds.), *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, pp. 31–40 Cambridge, MA. MIT Press.

Smolin, L. (1997). *The Life of the Cosmos*. Oxford University Press, Oxford.

Sober, E. (1994). *Conceptual Issues in Evolutionary Biology* (2nd edition). A Bradford Book, MIT Press.

Sober, E. (1996). Learning from functionalism — Prospects for strong artificial life. In Boden, M. A. (Ed.), *The Philosophy of Artificial Life*. Oxford University Press, Oxford.

Steels, L. (1994). The artificial life roots of artificial intelligence. *Artificial Life*, *1*, 75–110.

Taylor, C., & Jefferson, D. (1994). AL as a tool for biological enquiry. *Artificial Life*, *1*, 1–13.

Thelen, E., & Smith, L. B. (1994). *A dynamics systems approach to the development of cognition and action*. Bradford Books/MIT Press, Cambridge, MA.

Thompson, A. (1997). Silicon evolution. In Koza, J. R., Goldberg, D. E., Fogel, D. B., & Riolo, R. L. (Eds.), *Proceedings of Genetic Programming 1996 (GP96)*, pp. 444–452 Cambridge, MA. MIT Press.

Todd, P. M., & Miller, G. F. (1991a). Exploring adaptive agency II: Simulating the evolution of associative learning. In Meyer, J.-A., & Wilson, S. W. (Eds.), *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behaviour* Cambridge, MA. MIT Press/Bradford Books.

Todd, P. M., & Miller, G. F. (1991b). Exploring adaptive agency III: Simulating the evolution of habituation and sensitization. In Schwefel, H.-P., & Maenner, R. (Eds.), *Proceedings of the First International Conference on Parallel Problem Solving from Nature* Berlin. Springer-Verlag.

Todd, P. M., & Miller, G. F. (1993). Parental guidance suggested: How parental imprinting evolves through sexual selection as an adaptive learning mechanism. *Adaptive Behavior*, *2*(1), 5–47.

Trivers, R. L. (1972). Parental investment and sexual selection. In Campbell, B. (Ed.), *Sexual Selection and the Descent of Man 1871–1971*, pp. 136–179. Aldine, Chicago.

Turing, A. (1952). The chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond., Series* B, *237*, 37–72.

Veblen, T. (1898). Why is economics not an evolutionary science?. In Lerner, M. (Ed.), *The Portable Veblen*, pp. 215–240. Viking Press, New York. Collection published 1948.

Veblen, T. (1899). The theory of the leisure class. In Lerner, M. (Ed.), *The Portable Veblen*, pp. 53–214. Viking Press, New York. Collection published 1948.

Vega-Redondo, F., & Hasson, O. (1993). A game-theoretic model of predator-prey signaling. *Journal of Theoretical Biology*, *162*, 309–319.

von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior* (Princeton edition). Princeton University Press.

von Neumann, J., & Burks, A. W. (1966). *Theory of Self-Reproducing Automata*. University of Illinois Press, Urbana, IL.

Walter, W. G. (1963). *The Living Brain*. W. W. Norton, New York.

Webb, B. (1994). Robotic experiments in cricket phonotaxis. In Cliff, D., Husbands, P., Meyer, J.-A., & Wilson, S. W. (Eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on the Simulation of Adaptive Behaviour*, pp. 45–54. MIT Press.

Werner, G. M., & Dyer, M. G. (1991). Evolution of communication in artificial organisms. In Langton, C. G., Taylor, C., Farmer, J. D., & Rasmussen, S. (Eds.), *Artificial Life II — SFI Studies in the Sciences of Complexity*, Vol. X, pp. 659–687 Redwood City, California. Addison-Wesley.

Wertheim, M. (1997). *Pythagoras' Trousers: God, Physics and the Gender Wars*. Fourth Estate, London.

West-Eberhard, M. J. (1979). Sexual selection, social competition and evolution. *Proceedings of the American Philosophy Society*, *123*, 222–234.

Wheeler, M. (1995). Escaping from the Cartesian mind-set: Heidegger and artificial life. In Morán, F., Moreno, A., Morelo, J. J., & Chacón, P. (Eds.), *Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life*, pp. 65–76. Springer.

Wheeler, M., & de Bourcier, P. (1995). How not to murder your neighbour: Using Synthetic Behavioural Ecology to study aggressive signalling. *Adaptive Behavior*, *3*(3), 273–309.

Wiley, R. H. (1983). The evolution of communication: Information and manipulation. In Halliday, T. R., & Slater, P. J. B. (Eds.), *Communication*, Vol. 2 of *Animal Behaviour*, pp. 156–189. Blackwell, Oxford.

Williams, G. C. (1966). *Adaptation and Natural Selection*. Princeton University Press, Princeton.

Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. Harvard University Press.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. Ablex, Norwood.

Yachi, S. (1995). How can honest signalling evolve? The role of handicap principle. *Proc. R. Soc. Lond., Series* B, *262*, 283–288.

Young, R. M. (1985). *Darwin's Metaphor: Nature's Place in Victorian Culture.* Cambridge University Press, Cambridge.

Zahavi, A. (1975). Mate selection — A selection for a handicap. *Journal of Theoretical Biology*, *53*, 205–214.

Zahavi, A. (1977a). The cost of honesty (further remarks on the handicap principle). *Journal of Theoretical Biology*, *67*, 603–605.

Zahavi, A. (1977b). Reliability in communication systems and the evolution of altruism. In Stonehouse, B., & Perrins, C. (Eds.), *Evolutionary Ecology*, pp. 235–239. Macmillan, London.

Zahavi, A. (1980). Ritualisation and the evolution of movement signals. *Behaviour*, *72*, 77–81.

Zahavi, A. (1982). The pattern of vocal signals and the information they convey. *Behaviour*, *80*, 1–8.

Zahavi, A. (1987). The theory of signal selection and some of its implications. In Delfino, V. P. (Ed.), *International Symposium of Biological Evolution*, pp. 305–327 Bari. Adriatica Editrice.

Zahavi, A. (1991). On the definition of sexual selection, Fisher's model, and the evolution of waste and of signals in general. *Animal Behaviour*, *42*, 501–503.

Zahavi, A. (1993). The fallacy of conventional signalling. *Phil. Trans. R. Soc. Lond., Series* B, *340*, 227–230.