UNIVERSITY OF
Southampton

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Physical Science
School of Electronics and Computer Science

The Interaction Between Lifetime Learning and Evolution

by

David Prosser

A thesis for the degree of
Doctor of Philosophy

May 2022

Research Thesis: Declaration of Authorship

Print name: David John Prosser

Title of thesis: The Interaction Between Lifetime Learning and Evolution

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research. I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. No part of this work has been published elsewhere.

Signature:

Date: 1st May 2022

University of Southampton

ABSTRACT

Faculty of Engineering and Physical Science

Doctor of Philosophy

The Interaction Between Lifetime Learning and Evolution

by David Prosser

The impact of learning on evolution has been subject to significant debate and analysis for over a century. Much progress has been made in explaining how learning, as a non-inherited, highly adaptive form of plasticity, affects evolutionary trajectories with the potential to become genetically assimilated via the Baldwin Effect. To date, most computational and mathematical models devoted to understanding the effect of learning on evolution only focus on how environmental input impacts the expression of a small number of independent traits. However, West-Eberhard suggests learning mediates selection pressures over many phenotypic traits thereby driving genetic correlations; these correlations, in turn, may increase the effectiveness of the learning, which with circular reinforcement, increases the effectiveness of genetic evolution. Here this concept is extended so that learning has the capacity to guide the evolution of linkages between innate behaviours leading to genetic assimilation of that learning through canalisation. This feedback is bi-directional: whilst learning improves the fitness signal to evolution, correlations between innate behaviours channel learning to enable the discovery of optimal phenotypes that would not normally be found by evolution alone. This dynamic is explored through novel computational models of two different causal paths between learning and evolution; one where learning has an indirect effect on the innate behavioural traits by changing the experienced environment and one where learning acts directly on the expression of innate behaviours. Both these models show that guided by learning, the evolution of correlations between innate behaviours produce high-fitness behavioural phenotypes faster and more consistently than without learning. Further, it is shown that the introduction of linkage between innate behaviours can relax the conditions under which learning can become genetically assimilated. All this work supports the emerging view of the phenotype as a vital actor in the process of evolution.

Acknowledgements

Thanks and appreciation to my supervisors Richard Watson and Markus Brede for their time, guidance, patience and good humour. Special thanks go to Christoph Thies, Jamie Caldwell and Frederick Nash for the time they have given to understanding and discussing this work. I would also like to thank Kostas Kouvaris, Miguel Brun-Usan, Laura Mears and Alfredo Rago for invaluable discussions and insights. The author also acknowledges the use of the IRIDIS High-Performance Computing Facility and associated support services at the University of Southampton.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The impact of in-lifetime learning on evolutionary processes has been debated since the time of Baldwin [6]. Recent discussion has framed learning as a form of plasticity that, with a few exceptions pertaining to complex species such as humans, is highly adaptive [129]. While some argue that, until recently, plasticity has been a neglected area of study [113], it is an accepted concept and is widely observed across multiple taxa. Examples range from the leaf growth response to reduced sunlight [112] to the morphological changes observed in the infamous walking two-legged goat [108]. Whilst these are examples of the environment changing the expression of physical traits, behaviour is also a form of phenotypic plasticity.

Innate behaviours, like developmental plasticity, have a strong genetic basis; the somatic cells controlling behaviours are products of phenotypic development guided by the individual's genetic code. Crucially, an innate behavioural response to stimuli is invariant over an organism's lifetime. Therefore, even when governed by a complex central nervous system[1], an innate behaviour could still be viewed as a simple, consistent response to the environmental trigger, just as a plant's response to lack of moisture is to divert resources to grow more roots [113]. However, learning has the potential to improve the response to complex environmental cues over the individual organism's lifetime (a temporally varying response), and thereby alter the effective fitness of an individual in a way that is genetically constrained but not fully genetically determined. Consequently, learnt behaviours have scope to impact either the speed or trajectory of evolution (or both) in a way that is different from both developmental plasticity and innate behaviours.

Moreover, innate behaviours and learning can impact evolution through a variety of causal routes including habitat selection [60], perturbational niche construction [60,

[1]There is evidence that innate behaviours do not require a central nervous system. For example, bacteria can exhibit goal seeking behaviour such as chemotaxis towards amino-acids [11]. Ginsberg & Jablonka [42] go as far to suggest that learning is exhibited in single-celled organisms such as *Stentor*

110], sexual selection [118] and non-genetic inheritance [2, 7]. This creates a complex mesh of potential effects of learning and some of the work of this thesis is dedicated to understanding the different modes of interaction.

As with other forms of plasticity, behaviours learnt within an individual organism's lifetime cannot be subject to direct genetic transmission between generations. If, as many ethologists believe[2], one assumes some innate behaviours were learnt by previous generations, the evolution of congenitally expressed behaviours is difficult to explain in terms of simple Darwinian incremental selection, variation and inheritance and therefore seems to run contrary to the Modern Synthesis [45]. Alternative mechanisms do exist for the inter-generational transmission of learning, such as parental teaching; but these mechanisms require observational learning that, whilst seen in a variety of species from humans to types of fish, is not a universal ability [15] and it may be viewed as the exception rather than the norm of learning across taxa [129]. This thesis focuses on exploring how learning can alter the course of evolution in the absence of social learning.

The Baldwin Effect [6], along with the allied concepts of genetic assimilation and genetic accommodation, describes a mechanism by which the plastic phenotypic response to extreme or varying environments provided by individual learning causes a shift in selection gradient, so that individuals who can plastically gain increased fitness in the changed environment, rapidly become more prevalent in a population. Over an evolutionary timescale, individuals that gain an adaptive plastic response to the new environment more easily prevail so that eventually the response will become a genetically determined trait. The incorporation of plastic response into the genome, can either be through selection on existing variation or mutation of the genome, including canalising the genotype through pleiotropic effects. This phenomenon has phenotypic adaptions as the lead and genes as the followers [129], so what was previously environmentally induced becomes *'constitutively produced'* [94].

Most empirical experimentation examining genetic assimilation in support of the argument for the Baldwin Effect has focused on developmental plasticity, the first and classic study being Waddington's artificial selection experiments [120] on *Drosophila* wings. It is worth noting that although the Baldwin Effect has been offered as an explanation for behavioural traits in many different species, there is comparatively little experimental evidence for a Baldwinian mechanism in support of these claims [48]. However, some experiments have provided evidence for learnt behaviours changing evolutionary outcomes. For example, in their study of the interaction between learning and innate behaviour, Mery & Kawecki's [78] experimentally demonstrated that *Drosophila Melanogaster's* innate preferences for laying eggs on pineapple, as opposed to the alternative of oviposition on oranges, increased more when previous generations

[2]For example, Tierney [116] and Price et al [96]; see Section 2.1.2 for more detail.

had also been conditioned to avoid oranges[3]. This effect may be due to assimilation but as Hayes, Charter & Dywer [48] point out, with Mery & Kawecki's experiment it is difficult to tell what process is increasing the preference; it could be an improvement to core cognitive function or peripheral mechanisms such as sensory systems. With this lack of extensive real-world evidence for the processes underpinning the Baldwin Effect, there has been a heavy reliance on analytical models and computational simulations to inform the debate about the action of learning on evolution. This thesis carries on this tradition by developing new models to explore how learnt behaviours can become innate via canalisation of the genotype through pleiotropic and epistatic effects.

Whether adaptive phenotypic plasticity, including learning, speeds or slows evolutionary processes has been hotly contested, with arguments, showing potential for plasticity to either hide the genome from the forces of selection, a *'hiding effect'* [73] or accelerate evolution; the Baldwin Expediting Effect [3]. Since the classic Hinton & Nowlan model [49] of the 1980s, there has been a plethora of models examining plasticity and learning in the context of evolutionary processes, often offering contradictory results. For example, while Pereira [92] argues learning does not accelerate evolution, in a similar model Downing [29] puts forward a more nuanced argument where learning can accelerate evolution if there is a simple phenotype/genotype map but has an inhibiting effect where there is an indirect development mechanism. Other computational and mathematical models, in support of this debate, have been constructed to specifically examine the impact of learning on evolution [12, 34, 74, 90] and others are designed to more generally explore the action of plasticity in core evolutionary processes [3, 4, 8, 10, 36, 40, 63, 101]. Much of the discussion has been primarily focused on the shape of the fitness landscape governing selection and this is extensively reviewed in Section 2.4. In addition, there have been contributions to this debate from the Artificial Life (ALife) community showing that learning is sometimes beneficial to evolution but not under all conditions [1, 37, 65, 76, 87, 93]; although these models tend to lack explanatory power. This work confirms that learning has the potential to increase the rate of evolution but, dependent on the causal route by which the learning outcome influences evolution and other prevailing conditions, can also remove selective pressure inhibiting evolutionary progress.

As well as the potential to change the speed of evolution, there is scope for plasticity in general, and learning in particular, to change evolutionary trajectories to attain evolutionary outcomes that would not normally be achievable; potentially reaching a higher population fitness. This Baldwin Optimizing Effect [133] primarily arises when phenotypic plasticity allows the population, that would, without plasticity, be stuck at a local optimum, to move to higher optima [80]. This work shows that learning has the potential to alter the effective fitness landscape or change the way the landscape is explored

[3]It was found that the opposite was true when Drosophila were conditioned to avoid pineapples with the non-conditioned population preferring oranges more quickly than the conditioned population. This makes it difficult to draw definitive conclusions from Mery & Kawecki results.

in a way that allows the discovery of optimum phenotypes that would otherwise not be reliably found without learning.

Whilst, as touched on above, much of the debate about the interaction of learning and evolution focuses on the outcome (i.e. an acceleration or optimizing effect) there has been less discussion of the conditions under which that outcome occurs. Firstly, there is uncertainty whether a cost of learning is needed as a selective pressure to drive assimilation of the learnt behaviour into the genotype. Whilst most models of learning and evolution need a cost of learning for genetic assimilation, others do not; for example, Borenstein et al's random walk model [12]. Secondly, is the issue of what Mayley [74] termed *'neighbourhood correlation'*. This challenge arises where plastic changes in the phenotype cannot be easily achieved by incremental mutations to the genotype, thereby inhibiting the incorporation of the learning into the genetic code. Thirdly, in the absence of transmission of learning across generations, it is often argued that the same learning would need to be consistently repeated over a large number of generations to provide a stable target for the genetics to assimilate to [18,129]. To support the examination of the necessary conditions for assimilation of learned behaviours, extensions to two of the aforementioned models have been implemented: Borenstein's [12] random walk model of evolution and learning in a multi-peaked landscape is adapted to include a cost of learning to investigate whether it facilitates or inhibits evolutionary outcomes and Mayley's [74] model of neighbourhood correlation has been revised to enable a more comprehensive search of phenotype space to explore whether this ameliorates the neighbourhood correlation issue.

Traditionally, models of learning and evolution either have innate behaviours replacing learnt behaviours (as in the Hinton & Nolan model [49] and its many derivatives) or the behaviours comprising of a fixed innate component and a varying learnt component (for example Papaj's model [90]). This latter approach reflects the view of phenotypic plasticity as a simple *'norm of reaction'*. Therefore many analytical models of learning and evolution adequately use single or dual-trait representations of the genome. However, if an individual's innate behaviour is integrating multiple stimuli, the response to environmental stimuli is likely to produce complex, non-linear and, with learning, potentially temporally-varying reaction norms [131]. A hypothesis that may offer a richer, multi-trait description of the interaction between learning and evolution is West-Eberhard's [129,130] suggestion that learning is a form of plasticity that can act to form correlations between physical traits: learned behaviours improve fitness and therefore mediate the forces of selection simultaneously across multiple traits leading to correlations between these traits. These correlated traits then enhance the effectiveness of the recurrent learning in future generations; leading to a virtuous feedback loop. An example of this regarding physical traits is Corning's [21] giraffe example, learning to reach higher for leaves will reward a longer neck, and longer necks reward reaching higher. This hypothesis is discussed further in Section 2.7.4.

In this thesis, West-Eberhard's concept is extended to encompass the creation of correlations between innate behaviours by a mutual interaction between the rapid mediating effects of learning and the slower evolution of correlations between genes controlling those innate behaviours. Where correlations between multiple traits are evolving through the fitness screening effects of the learnt behaviours, the search-space of combinations of innate behaviours is narrowed in a way that allows solutions to be found that would not normally be encountered without the learning. In the opposite direction, the evolved correlations between innate behaviours work to constrain learning by making some responses much more rewarding than others. This is the genetic assimilation of learnt behaviours through canalisation. Using this mechanistic explanation as a framework, this work aims to show that learning, working in concert with innate behaviours, not only accelerates the evolutionary process but also achieves higher fitnesses that learning or evolution cannot attain in isolation. Unlike most models of the Baldwin Effect (for example Hinton & Nowlan), learnt and innate behaviours are not battling for control of the phenotype. It, therefore, demonstrates the potential for a significantly different explanation for the action of the Baldwin Effect.

This canalisation model of the Baldwin Effect is explored by considering two different causal routes for the interaction between learning and evolution. In the first model, learning is represented as directly testing different combinations of innate behaviours, where evolved correlations between innate behaviours make them more likely to be trialled together with high-fitness combinations retained. This Correlated Behaviours model shows a significant acceleration and optimizing effect without the need for the same learning to be repeated over multiple generations. The second model explores an indirect mechanisms for learning to influence the evolution of innate behaviour where learning can alter the effective environment [110] experienced by an individual so that the selective pressures across multiple behavioural traits are changed. This Environment Selection model also shows an expediting effect without the need for a cost of learning to drive genetic assimilation and without the need for the learning to directly influence behavioural trait values. This indirect route for the assimilation does not have learning directly searching for the correct phenotypic configuration, but rather has learning as an amplifier of the fitness signal between variation and selection.

Both models simulate assimilation of the learning through canalisation and provide important insights into how assimilation can occur where there is a complex mapping between genotype and phenotype. These models of learning and evolution are preceded by a model that is used to explore a correlation model of plasticity using the simpler scenario of environmentally sensitive developmental plasticity (ESDP). As well as demonstrating that plasticity can accelerate evolution based on the plasticity guiding the canalisation of the phenotype, this model also introduces some core modelling representations and assumptions as well as investigating some unexpected dynamics.

This work presents the phenotype as a vital actor in the evolutionary process. This view is contentious as it has been argued that adaptive plasticity is just a product of evolutionary processes and therefore it is still evolution that is doing all the work [127]. However, the different qualities of learning explored in this thesis set it apart from other forms of plasticity, decoupling the phenotypic selection from the genetic response and, therefore, provides some support for a phenotype-first explanation for the evolution of individual behaviours.

The key contributions of this work are:

1. An exploration that elucidates the necessary conditions under which learnt behaviours are normally genetically assimilated.
2. A hypothesis that there is a two-way feedback mechanism between learning and evolution, where learning can guide the evolution of genetic correlations between innate behaviours and those correlations channel what is learnt.
3. A demonstration of a canalisation model of the Baldwin Effect where, once found by adaptive plasticity (morphological or behavioural), the alternative expression of the phenotype can be assimilated via canalisation due to epistatic and pleiotropic effects.

The claims of this thesis are:

1. Canalisation of the genotype can be guided by learning (a) directly, when the action of learning alters the expression of innate behaviours or (b) indirectly, when learning alters the experienced environment.
2. Direct two-way feedback between learning and evolution can change the movement of the phenotype in the fitness landscape in a way that allows rapid discovery of globally optimal fitness phenotypes that are unlikely to be reliably found by evolution alone.
3. The conditions normally associated with the genetic assimilation of learning are relaxed when that assimilation is achieved by the genetic linkage between innate behaviours.

Chapter 2

Foundations

The effect of learning and innate behaviours on evolutionary outcomes is a complex topic with significant disagreement on both terminology and substance. As Frank [35] puts it:

> *"However, the jargon from those theories is thick: the Baldwin effect, genetic assimilation, reaction norms, hopeful monsters, niche construction, and environmentally induced evolution. Each variant theory invoked special environmental conditions, developmental processes, and interactions with genetics. And each in its own way jousted with the ghost of Lamarck. A casual observer could be forgiven for steering clear of the whole mess."*

Consequently, understanding and untangling the concepts associated with behavioural and developmental plasticity is an important first step to developing a more holistic model of how learning changes evolutionary processes.

In this chapter, the differences between alternative forms of plasticity are examined and how these differences impact evolutionary outcomes assessed. This informs how models of learning and evolution need to differ from models of developmental plasticity and evolution. We then consider the different mechanisms by which behaviours, both learnt and innate, expressed during an individual's lifetime can influence evolutionary outcomes. This provides a framework for the different causal paths by which learning can alter evolution outcomes with a subset of these routes tested by the models presented in this thesis. How learning can be incorporated into the genome to become inherited behaviours - without resort to Lamarckian mechanisms - is then discussed; this effect being fundamental to learning having a lasting impact on evolutionary outcomes. Finally, we look at how the effects of learning acting over multiple traits can cause mutual feedback between learning and evolution that may offer an alternative explanation for the assimilation of learning.

2.1 Behaviour as a Form of Phenotypic Plasticity

2.1.1 Developmental Plasticity, Innate Behaviour and Learning

Phenotypic plasticity refers to the ability of a single genotype to express multiple phenotypes due to environmental influences. Whilst it is most commonly associated with the alternative expression of physical traits during the development of the phenotype, it can equally apply to behaviours, which are generally viewed to be either innate (instinctual) or learnt (acquired). Consequently, phenotypic plasticity can be divided into two broad categories: developmental plasticity and behavioural plasticity. Developmental plasticity is most commonly described as the *"ability of genetically identical organisms to* ***develop*** *(my emphasis) different phenotypes in response to different environmental conditions"* [25]. Behavioural plasticity can either be learnt behaviours or innate behaviours or a combination of the two. Innate behaviours are inherited responses to environmental stimuli that do not vary over time and learnt behaviours are discovered during a lifetime and alter contingent on an individual's previous experience. Both innate and learnt behaviours are environmentally sensitive [46], in that they respond to environmental cues (and internal states, such as hunger). For example, a bird may innately start its migration based on the environmental trigger of the length of the day [75], whereas a honey bee will learn to associate a specific colour with a source of nectar and another colour with a source of pollen [85]; the behaviour of visiting a certain colour of flower triggered by the need to collect pollen or nectar.

Other categorisations of behavioural plasticity have been offered: Snell-Rood [109] distinguishes between innate behaviour and learnt behaviour using slightly different terms: *Developmental behavioural plasticity* encompasses any change to the nervous system as a result of experience but also includes any physical change caused by that behaviour (e.g. effect on muscles) whereas *activational behavioural plasticity* is akin to innate behaviours in that the behaviour is activated by environmental cues. Whilst the above nuances of terminology exist, for ease of reference, we shall refer to developmental plasticity as the expression of physical traits influenced by the environment (rather than behavioural traits) which whilst may be most prevalent during development can occur throughout an organism's lifetime. This aligns with the common usage of the term.

Much of the published thinking on the interaction between evolution and learning, acknowledges innate behaviour and learning as a form of phenotypic plasticity but too often are not specific about the characteristics of the plasticity being considered[1]. Careful consideration of learning, innate behaviour, and developmental plasticity reveals that the different types of phenotypic plasticity have properties that have the scope to

[1]For example Diogo [27] very seldom uses the word learning in his extensive book and does not attempt to distinguish between innate and learnt behaviours

change evolutionary processes in different ways. Therefore, for convenience, we distinguish between innate and learnt behaviours as distinct forms of behavioural plasticity. The substantive differences between forms of plasticity underpin differing assumptions in previous analytic models, as well as those that underpin the new simulations presented in this work, and are discussed in the following sections.

2.1.2 Learning as a Regulator of Innate Behaviours

The previous section of this chapter discusses the categorisation of different forms of plasticity as if these are entirely separate phenomenon. However, it seems likely that rather than being discrete phenomena, each form of plasticity has the potential to interact with the others. For example, Mery & Burns [77] are clear that learning and innate behaviour are not binary options, with an organism's behaviour most likely a result of a combination of inherited and acquired actions based on lifetime experience; innate behaviours can be adapted by learning and may provide a linkage as to which environmental cues are important for which behaviours. This definition is supported by Konrad Lorenz's [69] observation that learnt behaviours and innate behaviours are not necessarily opposing concepts, with the responses that each individual exhibits being a combination of both instinctive and learnt responses. Behaviours have been observed in nature as a combination of social innovation and congenitally expressed behaviour. For example, the modification and use of tools made from twigs or cactus splines by Galapagos Island Woodpecker Finches (*Catospiza pallida*), have been shown to be a combination of a 'learning disposition' and trial-and-error learning [115]. As West-Eberhard neatly summarises: *"Learning is a regulatory mechanism that influences the expression of certain behavioural traits and the use of certain morphological ones"*. West-Eberhard concludes that the regulatory mechanism provided by learning leads to the recurrence of particular behaviours providing high fitness benefits. Mery & Burns [77] go on to speculate that the balance between behavioural plasticity and learning will be dependent on the degree of environmental heterogeneity. Learning will be selectively favoured where the environment is temporally or spatially inconsistent within a generation, as the rapid adaptation offered by learning will have a significant fitness benefit, whereas innate behavioural plasticity will be favoured where change is occurring gradually over multiple generations as evolution has time to adapt and the costs of learning will be avoided.

The premise that learning regulates innate behaviours and the implied hierarchy in Figure 2.3 suggests that learning is a more highly evolved characteristic than innate behaviours. This then raises a key question for any model of instinctual behaviours and learning: do innate behaviours become increasingly more sensitive to multiple repeated stimuli and make a transition to being learnt behaviours, or even does learning evolve separately to innate behaviours? Alternatively, do learnt behaviours become innate with learning being replaced by, potentially less costly innate behaviours through some mechanism that facilitates assimilation of learning into the genome? Interesting, Tierney [116] presents the ethologist's conventional wisdom as being that learning requires a more complex central nervous system than innate behaviour and therefore learning is a more highly evolved characteristic. Lorenz [69] appears aligned with this view when he points out that unicellular organisms that do not learn can still orient themselves and have the *'insight'* as to their direction: these behaviours are innate without first being learnt. This would suggest that, in some cases, innate behaviours evolve before learning. This view is supported by Dridi & Lehmann [32] citing Kerr [57], who suggested that *"learning is more likely to evolve on top of innate behavioural plasticity"*. A recent simulation of the evolution of path-following behaviour in an open-ended environment by Pontes et al [95], where computer instructions controlling the phenotype are evolved, supports Kerr's view: the results of this simulation showed that instinctive behaviours were evolved and then adapted to become learnt behaviours. This suggests that simpler behaviours are a necessary precursor to learnt behaviours and *"simple reflexive behaviours provide the building blocks for learning to arise"* [95].

However, Tierney [116] uses insights from neurobiology and nervous systems development to challenge this conventional view and hypothesizes that *"learned behavioural adaptations may commonly precede innate forms of the same behaviours"*. In support of this, she also hypothesizes that *"behavioural flexibility is phylogenetically primitive ..."*. Price et al. [96] also suggest that learning behaviour first appears through pure trial-and-error and then is spread through the population. Therefore, they reason that the innate behaviour observed today must result from a previous period of genetic accommodation, i.e. learnt behaviours appear in advance of innate behaviours. West-Eberhard [129] echoes Price's view by suggesting that recurring learnt behaviours could become genetically accommodated.

It seems plausible that both these views are correct, with some learnt behaviours becoming innate and some innate behaviours becoming increasingly sensitive to previous stimulation to become learnt. Most models of learning and evolution assume that learnt behaviours can be genetically assimilated to become innate behaviours (see Section 2.3). In line with West-Eberhard's [129] observations, the models in this thesis are built on the premise that learning utilises a repertoire of innate behaviours and that learnt behaviours can be genetically assimilated through different causal pathways to become innate. This is explored more fully in Section 4.2.

Whilst, the above establishes that it is difficult to categorise any given behaviour as either purely innate or learnt, the characteristics of the two different forms of behaviour, as well as how they relate to developmental plasticity, for convenience in the following sections are considered in their pure forms.

2.1.3 Behaviours are Normally Adaptive

Plasticity can have different impacts on phenotypic fitness. As discussed by Ghalambor [38], there are two potential plastic responses to environmental conditions: adaptive, where the plasticity increases the average phenotypic fitness within a new or changed environment and non-adaptive where fitness decreases in a novel environment. For developmental plasticity, it is easy to understand that plasticity can be adaptive, as per the leaf size example shown in Figure 2.1. Maladaptive developmental plasticity often occurs when the plasticity cue in the novel environment is not reliable. For example, fresh-water snails plastically respond to molluscivorous sunfish predators by forming more rotund shells and reducing growth; both responses increasing survival. However, this developmental response is also triggered in the presence of non-molluscivorous sunfish and therefore is maladaptive in non-molluscivorous environments [64].

Like developmental plasticity, genetically determined innate behaviours may be non-adaptive with novel stimuli. A goose's instinctual egg retrieval response to eggs that have rolled outside the nest is particularly well triggered by beer bottles [44]. In a beer-bottle rich environment, this non-adaptive innate behaviour is likely to be quickly removed by selection. However, for learning, it is commonly assumed that plasticity is adaptive. As argued by West-Eberhard [129], while there may be maladaptive learning with more advanced species, it is unlikely to be a common occurrence for the majority of phyla. In this thesis, we assume that learning is locally adaptive by which we mean that learning increases fitness towards local fitness peaks and that innate behavioural expression can, in theory, be both adaptive and non-adaptive, i.e. can move phenotypic fitness both towards and away from local fitness peaks.

2.1.4 Behaviours are Reversible

Dridi and Lehmann [32], citing Botero et al. [13], identify that developmental plasticity is usually considered irreversible within a lifetime whereas behaviour and learning are both reversible; the behavioural trait is only manifest in the presence of a certain environmental stimulus and, when the stimulus is not present, the behaviour is not exhibited. An example of irreversible developmental plasticity is the increase of leaf size in response to a reduction of sunlight [114]: the leaf cannot reduce in size once it has grown. Although developmental plasticity is usually thought of as irreversible, this permanency is not universal. For example, West-Eberhard [129] describes Cordero's work examining sex colouration, where some species show reversible colouration changes that are temperature-dependent. Indeed, Chenard & Duckworth [16] recently argued that morphological traits can also have a range of flexibility similar to that of behavioural plasticity. They use the examples of a snake's gut expanding after a large meal and the changes to skin tone after exposure to the sun as examples of reversible plasticity. Using the example of 'personality' traits such as aggression, they also argue that some behaviours have similar stability as morphological traits - consistently being exhibited during a lifetime. However, a personality trait usually manifests itself in response to a specific or set of specific stimuli, rather than constantly so feels different to a morphological trait such as leaf-size that is variant during development but is invariant after that.

Reversible plasticity is sometimes referred to as *activational plasticity* or *labile plasticity* [25]. The reversibility of behaviour is a key characteristic of why it has the potential to speed up evolution; it allows the within lifetime expression of different phenotypes (with different fitnesses) thereby exploring a fitness landscape that would take multiple generations to search relying on irreversible plasticity. However, as seen in Chapter 5, irreversible developmental plasticity still has the potential to speed evolution by finding higher fitness phenotypes that can provide a better quality fitness signal to evolution.

2.1.5 Learning is Temporally Varying; Innate Behaviour is Not

The plastic response to an environmental factor is often illustrated with a simple norm of reaction where a trait is plotted against different environmental cues as shown in Figure 2.1. Fitness is often defined as a monotonic function over the set of environments. This representation reflects the static nature of developmental plasticity where there is a largely deterministic response to the environmental cue. However, is this representation consistent with behavioural plasticity?

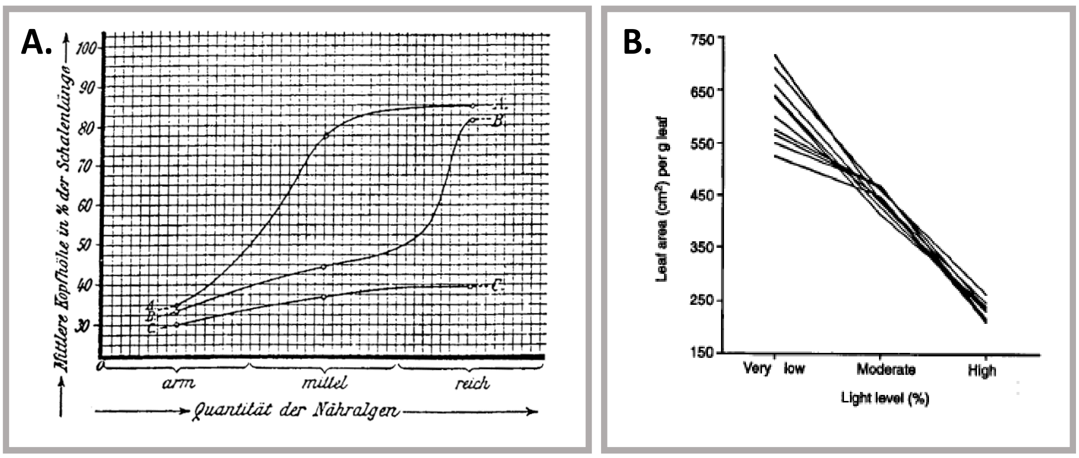


FIGURE 2.1: An illustration of two monotonic norms of reaction. **A.** From Woltereck [132] in 1909, norms of reaction for Daphnia head size found in different German lakes. **B.** Reproduced with permission from Sultan [112] in 1995, norms of reaction for leaf size by the environmental cue of sunlight

For innate behaviours, assuming consistent environmental stimuli, the response is still invariant over time as there is no influence of previous experience to vary the response. Consequently, innate behaviours can be thought of as akin to developmental plasticity with a response that is independent of the history of stimuli. However, Corning [21] states that individuals at different levels of complexity - from bacteria to chimpanzees - make behavioural choices and are capable of new responses to novel stimuli. From this, he concludes that the mapping between a series of environmental cues (experiences over a lifetime) and behavioural outcomes will be more complex for behavioural choices influenced by learning than those that are purely innate.

This suggests a highly dynamic norm of reaction with a state-dependent mapping between the response and the current stimuli, sensory acuity, previous stimuli, the effectiveness of recall and the ability to combine multiple inputs. Therefore, a crucial difference between a learnt behaviour, innate behaviour, and developmental plasticity is the dynamism of the reaction norm. Figure 2.2 uses the norm-of-reaction representation to help clarify the difference between innate behaviours and learnt behaviours in terms of the impact of environmental stimuli over time.

2.1.6 Population-level Innate Behaviours and Individual-level Learning

Plasticity has the potential to act at the level of the individual or the population. Developmental plasticity acts at the population level in that, assuming consistent environmental conditions for the entire population, the frequency of a plastic response will change from one generation to the next depending on the fitness of that response and how that fitness changes selection for the alleles determining the plasticity[2]. However, at what level the plasticity is acting is not so clear for innate behaviour and learning. Mery & Burns [77] contrast innate behavioural plasticity and learning by describing innate behavioural plasticity as something that is evolved at a population level over multiple generations with the response being dictated by the genetically determined response to the environmental input. In contrast, they categorise learning as something that happens within an individual over its lifetime and is therefore acting at the individual level. Mery & Burns appear to assume that for innate behaviours, the whole population is subject to the same environmental stimuli. If not, there is scope for innate behaviours to act at the sub-population or individual level.

[2] One reason that developmental plasticity may have a much more powerful effect than random genetic mutation is that the environmental change often happens to the entire population, simultaneously testing a range of plastically expressed phenotypes and, unlike random mutation, the phenotypic change has a consistent direction of expression across the population due to the change in environment [66]. However, it is also worth noting that a large standing genetic variation in the population would be tested by selection in the same way as developmental plasticity.

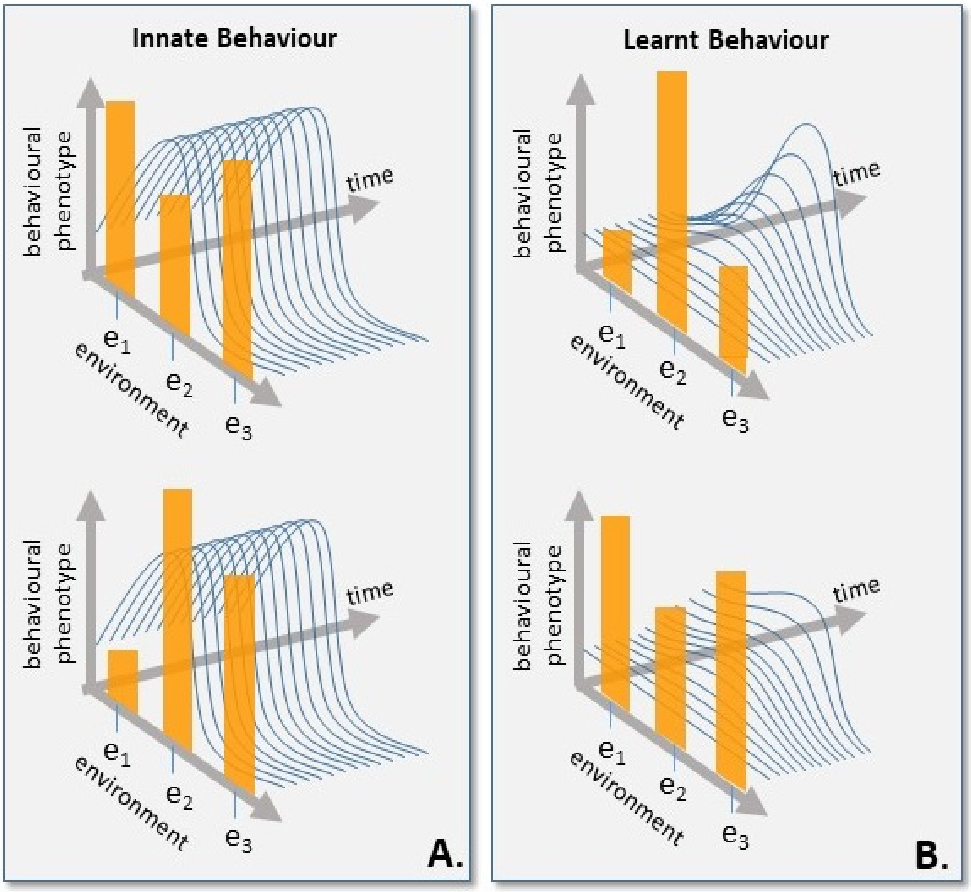


FIGURE 2.2: **A.** For innate behaviours, the norm-of-reaction of the *'behavioural phenotype'* [99] (depicted by blue lines) remains invariant over time regardless of the amount of time receiving a stimulus from each environment (relative exposure denoted for each environment by orange bars). **B.** For learnt behaviours, the norm of reaction (blue lines) is a function of the mix of current and previous environmental cues and therefore varies depending on an individual's history

Despite the difference between levels of action, both learned and innate behaviours have the potential to accelerate evolutionary outcomes. In common with developmental plasticity, innate behaviours are tested at the population level (assuming that the whole population is subject to similar environmental stimuli) and therefore, like developmental plasticity, simultaneously testing a range of genotypes (controlling behaviour), accelerating the movement of the population to fitter phenotypes. The reversibility of learning, as discussed in Section 2.1.4, combined with the ability to retain fit behaviours during a lifetime provides learning with the potential to create what West-Eberhard terms *'a fitness screening effect'* [129]. Through trial-and-error, this can increase fitness by testing multiple behaviours within a lifetime akin to evolution within a lifetime. A key topic for this thesis is, ignoring social transmission of learning such as mimicry, how can this learning at the individual level impact the population as a whole without resorting to the Lamarckian inheritance of acquired characteristics?

2.1.7 Genetic Determination of Plasticity

From Section 2.1.5, we can make some assumptions as to the extent to which the different forms of plasticity are determined by genetics and the environment. For developmental plasticity, the case seems straightforward: the plastic response is a direct product of genetics and environmental inputs. The environment can change the course of evolution, but the organism's plastic response is still genetically determined. The same can be argued for innate behaviours: the interaction between the genome and environment is more complex because there can be a range of responses within a lifetime - each with its own fitness consequence - based on a variety of environmental triggers. However, the response to the environmental trigger is still entirely genetically determined.

For learning, behavioural outcomes are not a direct product of genetic disposition and the current environment: the accumulation of experience including that from social interaction and soft inheritance mechanisms such as parental teaching, all make the link between an individual's genome and their behavioural outcomes much less direct. As Laland [61] points out, many evolutionary theorists now *"view phenotypes (and hence their environmental modification) as underdetermined by genes."*. However, regardless of the machinery of learning, the fitness benefits of acquired behaviours are constrained by the phenotypic traits that are a product of the genotype influenced by the environment. For example, an organism's ability to learn to fly is constrained by the extent to which it has rudimentary wings. Therefore, one can argue that learning is, at the very least, genetically constrained; given that learning requires biological machinery to support its action, as well as physical traits that constrain learning, it can also be viewed as genetically facilitated.

2.1.8 Behaviours Change Evolutionary Outcomes

As can be seen in Figure 2.3, there are significant differences in the characteristics between developmental plasticity, innate behaviours and learning. Consequently, different forms of plasticity will influence evolutionary outcomes in different ways.

Whilst developmental plasticity can influence evolutionary outcomes, the reversibility of both learnt and innate behaviours allows increased adaption to new or changing environments within the organism's lifetime. Like developmental plasticity, in general, innate behaviours will be tested across the whole population in a consistent direction with the potential to have a strong effect on evolutionary trajectories. Further, the genetically under-determined nature of learning detaches the individual from the evolutionary timescale potentially allowing high fitness behaviours to be discovered that would take many generations to achieve if relying on the random mutation of the genome. However, without some causal route for learnt behaviours to be transmitted

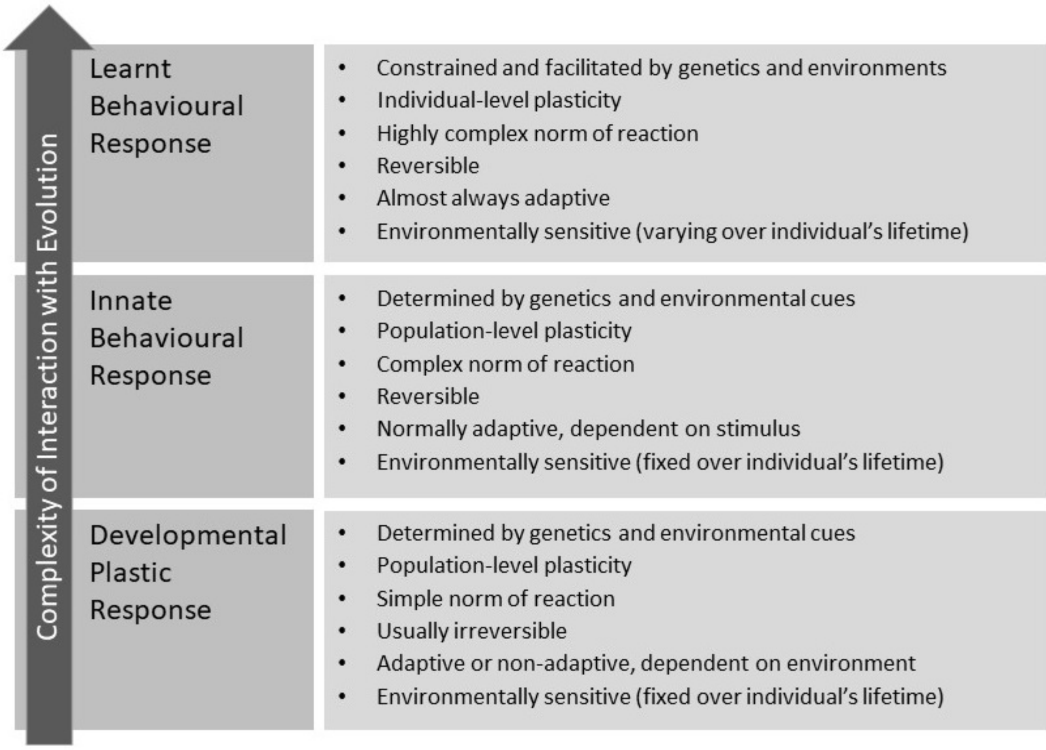


FIGURE 2.3: Summary of key differences between developmental plasticity, innate behaviours and learnt behaviours.

to the next generation, this learning will be lost when the individual dies and so the learning will have a limited lasting impact on evolutionary outcomes. The potential causal pathways for learning to impact evolution are examined in the next Section.

2.2 How Behaviour can Change Evolution

2.2.1 Behaviour and Evolutionary Causation

It is widely debated as to whether environmental influences on the phenotype can change evolutionary outcomes (see review in Levis & Pfennig [66]). The 'phenotype-first' or 'plasticity-first' view of evolution raises both philosophical questions and practical implications for evolutionary research. Whilst a detailed exploration of the 'phenotype-first' question is a thesis in its own right, the consideration of learning in terms of the ultimate and proximate causation helps crystallise the definition of the different ways in which learning can impact evolutionary outcomes.

In the seminal paper in 1961, Ernst Mayr [75] provided a distinction between two types of causation relevant to the study of evolution and biological systems. He defined 'proximate causation' as being concerned with the "what" of biological systems: the mechanisms by which the information in the DNA instructs cells to develop an organism; the domain of the functional biologist. Mayr contrasts this with 'ultimate causation', which is concerned with "how" DNA came to programme the organism through its evolution history; the domain of the evolutionary biologist. To illustrate this, he used the example of warbler migration, where there is an ecological cause (the bird must migrate or starve through lack of suitable insects), a genetic cause (the bird's genetics induces it to migrate), an intrinsic physiological cause (it is responding to environmental factors such as day length), and an extrinsic physiological cause (a sudden drop of temperature caused the bird to migrate on a particular day). Mayr argues that the latter two causes are proximate causes, whereas the first two causes are the ultimate causes. Crucially, proximate causes act at the individual level and ultimate causes *"on a particular DNA code of information with which every individual in every species is endowed" [75]* by which we might infer that ultimate causes act at the population level.

More recently, Laland et al [61] suggest that the distinction between proximate and ultimate causes, deeply embedded within the established frameworks of evolutionary theory, has inhibited progress in evolution thinking. More specifically, they argue that the separation between proximate and ultimate causes present a one-way causation path for evolution (from ultimate causation to proximate causation) and is artificial when considering widely observed effects in evolution such as habitat selection, niche construction, and mating preferences. The cornerstone of a new, emerging framework for understanding evolution is reciprocal causation, where *"phenotypic plasticity, which is extremely widespread in nature, can generate selection and thus precipitate evolutionary episodes"* both in terms of evolved traits being influenced by the environment through development and biasing of gene expression as well as production and preservation of novelty. A counter-argument to this is that the phenotypic plasticity that responds to environmental cues is genetically determined. The response to environmental stimuli (i.e. the proximal cause) is genetically programmed, be it a previously biased developmental response or, as in the case of the warbler, an innate response: selection on a variation of the inherited DNA has still done all the work. However, Laland [61] suggests that genes are just one piece of the picture of evolution because there is a complex mesh of selective and proximal forces potentially involving cross-species interactions, epigenetic and cultural inheritance. This complex web alongside selection acting at multiple levels is evidence that exogenetic factors varying over multiple generations are on an equal par with genetically inherited factors. Learnt responses can be viewed as a part of this complex mesh, especially where the learning can cross the generational boundary (e.g. parental teaching).

One important implication of the interaction between proximate and ultimate causation in the context of learning is the need for the plastic response to cause genetic change that fixes in the population. Laland's reciprocal causation concept appears to assume that outcomes associated with proximate causation consistently influence ultimate causation (see arrows in Fig 1 of [61]). With examples such as niche construction, this holds true: a dam constructed by a beaver will impact the selection pressures on future generations. However, depending on the type of learning, in the absence of genetic accommodation or assimilation, and ignoring social transmission, the behaviours learnt within a lifetime will be lost unless they alter the environment. In this context, the potential effects that individual behaviour and learning can have on evolutionary outcomes are described with examples in the following sections.

2.2.2 Behaviour Directly Affecting Selection

There can be a simple and direct effect of both learnt and innate behaviours on selection which in turn can alter evolutionary trajectories. Taking the example of the influence of physical traits, predator avoidance behaviours have a high impact on fecundity. These behaviours can then, via selection, affect physical traits: the learnt or innate behaviour of climbing trees to avoid predators creates selection pressure for shorter legs, whereas running-away behaviour creates selection pressure for longer legs [71].

2.2.3 Perturbational Niche Construction

Perturbational niche construction is commonly associated with the construction of artefacts, such as beavers' dams. Consequently, niche construction differs from habitat selection in that the environment is altered by the behaviour. A simple example provided by Snell-Rood & Steck being how burrowing behaviour modifies temperature variation: a 30cm deep burrow has significantly less variation than the surface temperature [110]. They also identify that there is scope for complex feedback mechanisms between the environment modifying traits and traits modifying the environment and its effects (through niche construction): this affecting selection for and against plasticity. They also point out that as well as decreasing the experienced environmental variability, behaviour can increase variability: neophilia (enthusiasm for the new) and exploratory behaviour will increase the experienced environment favouring selection for increased plasticity allowing survival across a more diverse range of environments.

2.2.4 Habitat Selection

Behaviour has the potential to change the environmental conditions that an individual in the population is exposed to by moving its location, changing its experienced environment. Habitat selection is one of the most basic forms of behaviour [47]. Sometimes termed *'relocational niche construction'* [60], a simple example of this is migration to a different climate changes the effective temperature variation on the phenotype, changing selection pressure on insulating fur.

Habitat selection is different to perturbational niche construction because, as Snell-Rood [110] points out, directly altering the environment enables the potential for *'ecological inheritance'* - assuming offspring stay in the same environment - whereas simply selecting a different environment without altering it does not. However, as Laland et al. state *"Strictly, everything that an organism does changes the environment to some degree."* [60]. As Laland observes, there is potential for a beaver's dung to have more of an effect on the environment than the beaver's dam [60]. In addition, the impact of the perturbation can either be felt by the individuals in the population enacting the behaviour or by a different species.

2.2.5 Mate Preference

Mating preference is a behaviour that has the scope to significantly influence evolutionary outcomes. In a review, Verzijden et al. [118] identify that sexual selection is observed to be learnt over a variety of species from insects to sheep; the classic example being sexual imprinting common to many species of birds. They identified that in sexual selection, the learning could take two forms: learned mate preference and learned traits under selection (Verzijden use the example of birdsong). Sexual selection is a broad topic and is not a causal route to evolutionary change that is further considered in this thesis.

2.2.6 Behaviours Changing Genetic Expression

It is worth noting that both environmental conditions and behaviour can cause non-genetic changes that are inherited from one generation to the next through mechanisms that alter the expression of genes [2]. These mechanisms include histone modification and DNA methylation and are intertwined with the genetic aspects of inheritance [2]. Since the response is a product of the environment and interacts with the expression of the organism's DNA, it can be viewed as a form of inherited plasticity. This form of plasticity is influenced by environmental cues received during development; an example discussed by Schwabl [104], cited by Ghalambor [39] being the maternal transmission of androgens in eggs altering the intensity of begging of chicks. Since, as can be seen in Schwabl's study, this 'developmental behavioural plasticity' (response to androgen level), acts as a mediator to environmentally stimulated behaviours and could be viewed as similar to morphological traits that constrain or facilitate behaviours. This view of epigenetic factors being a mediator to the environment, as stated by Baugh & Day [7], has this epigenetic inheritance as passive. Baugh & Day [7], Adrian-Kalchhauser et al [2], Danchin et al [24] and others contend these factors should be viewed as a more active *'multigenerational plasticity'* where *'epigenes'* that encode information non-genetically are inherited, may be subject to variation and can be important to fitness. Importantly, Adrian-Kalchhauser's review provides examples of behaviour altering epigenes that then have effects on the next generation.

However, whilst it is possible for learning to cause biological changes within an individual that then can then affect the next generation (e.g. overeating in adults impacting their children's development [2]), it is not the learning outcome itself that is inherited and therefore there is little scope for assimilation of the learning that would be adaptive. Therefore, this causal route for the impact of learning on evolution is not considered further in this thesis.

2.2.7 Separating Causal Flows

Figure 2.4 summarises the causal flows between behaviour and evolution discussed in the previous sections. Panel A shows the simplest, no-plasticity case and then each panel below adds different routes by which developmental plasticity (in the case of Panel B) and behaviour (Panels C to E) can influence evolutionary outcomes.

One can also ask: How does the 'direct effect on trait' of running away from a predator differ from relocating to environments where there are no predators? An important difference is that habitat selection, unlike a direct effect on a trait, is likely to have an impact on a much broader range of traits unrelated to the behaviour. For example, whilst running away from predators may create a selection pressure on leg-length for

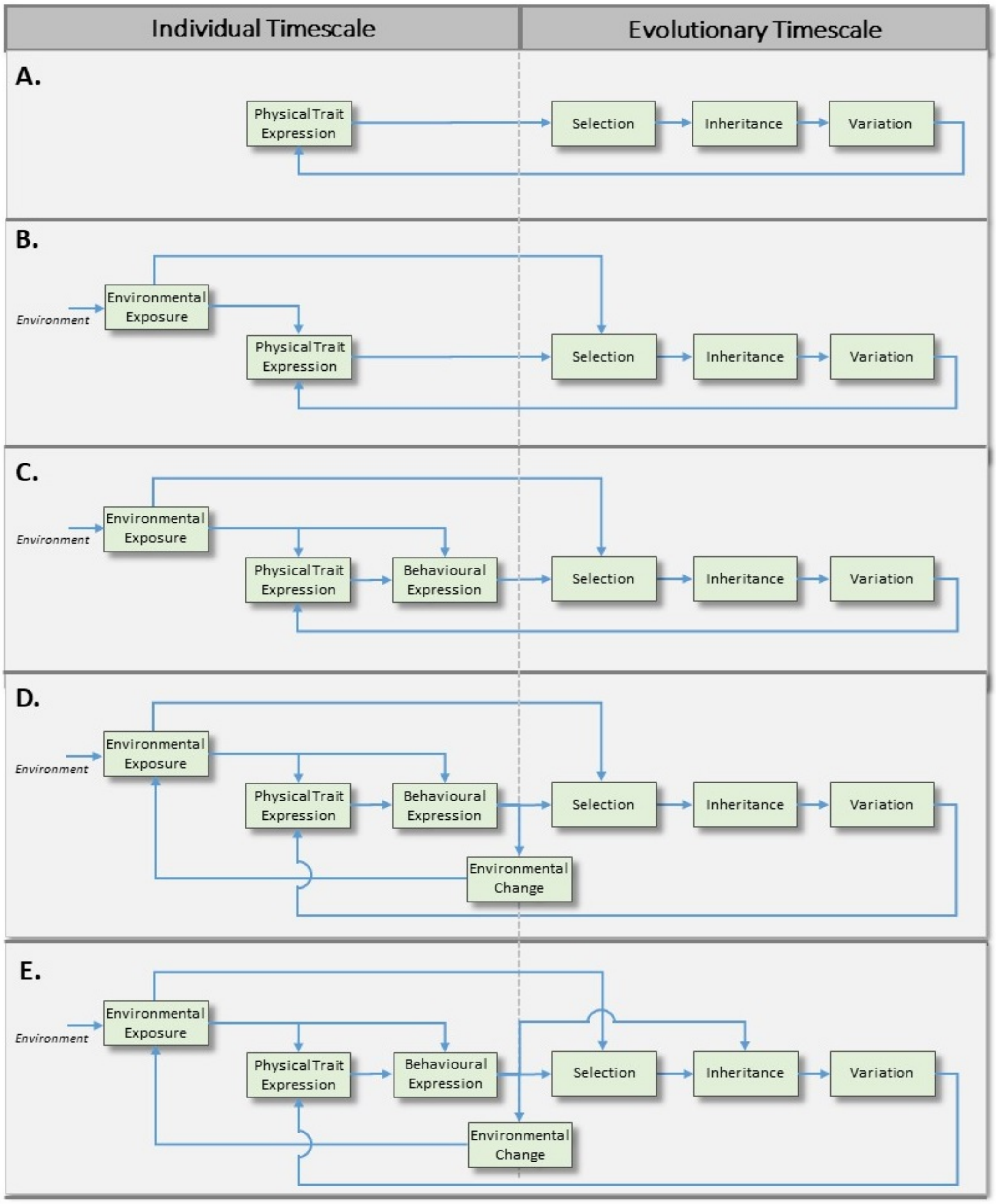


FIGURE 2.4: Behavioural expression can impact evolutionary outcomes through a variety of causal paths. **A.** Without any form of plasticity, trait expression is genetically determined and impervious to the environment. **B** Selection acts on physical traits that may be altered by the environment in which the phenotype develops (developmental plasticity). **C.** Behavioural expression is triggered by environmental stimuli and is constrained/facilitated by physical traits. The behavioural expression has fitness consequences that can directly alter the selection on the phenotype. **D.** Behaviour can indirectly temporarily or permanently change selection through habitat selection and/or perturbational niche construction which causes change to the environment to which the phenotype is exposed. **E.** As well as directly and indirectly impacting selection, behavioural expression can also impact inheritance through behaviours that alter the genetic expression in future generations.

speed of escape, migration may lead to selective pressure on leg-length and the amount of fur, as the new environment may have a different temperature profile.

As will be discussed extensively in this and subsequent chapters, learning ultimately influences evolution by changing the fitness of the phenotype in its experienced environment. This happens either by changing the phenotype physically or behaviourally or by the phenotype changing the experienced environment [110] (via physical changing of the environment or relocation). The core simulation work in this thesis, models three of the five causal flows in Figure 2.4 these being as depicted in Panels B, C and D. In the case where the experienced environment is changed, perturbational niche construction has not been included in the scope of this thesis and whilst this work considers these effects separately, as can be seen in the figure, many of the potential effects of behaviour on evolution are unlikely to be biologically separable.

2.3 The Baldwin Effect, Genetic Assimilation, and Genetic Accommodation

2.3.1 The Baldwin Effect

The traditional view of the core evolutionary process often labelled as the Modern Synthesis [45], is a blending of Darwin's theory of natural selection with Mendelian genetics: the gradual accumulation of small, heritable mutations of the genome, subjected to natural selection, leads to phenotypic change. This view is being increasingly challenged by proponents of phenotype-first [129] evolution who hypothesise that phenotypic plasticity guided by environmental factors is a key mediator of evolutionary change. When a species encounters or invades a new environment, plastic expression of traits enables the organism to survive and adapt quickly to the new environment allowing the genetic change, either through existing genetic variability or mutation, to 'catch-up' as the population is exposed to the new selection forces [128,129]. Essentially this view of evolution decouples, to some degree, the phenotypic selection from the genetic response [103]. Changes in population allele frequencies in the direction of an adaptive plastic response to a novel environment is known as the Baldwin Effect [107] and is often associated with selective forces replacing plastic responses with genetically determined responses. However, the Baldwin Effect is often conflated with the term genetic assimilation [22] which according to Crispo relies on canalisation against environmental perturbation (see Section 2.3.2). Whilst originally conceived by Baldwin as applying to learning, the concepts underpinning the Baldwin Effect can be applied more broadly to all forms of phenotypic plasticity [3, 22].

A basic mechanistic explanation for the Baldwin Effect is that selection under a new environmental condition, produces phenotypes that were not previously visible to selection under normal conditions and therefore these environmentally induced phenotypes and the alleles that facilitated the adaptive plastic response can be picked out by selection. This shifts the distribution of genotypes in the direction of the plasticity so that the phenotype can become reliably produced, and in some cases but not always, be produced in the absence of the environmental input. As West-Eberhard [129] points out, unlike genetic assimilation, the Baldwin Effect does not necessarily entail a reduced reliance on environmental input.

A simple three-phase process for environmentally-induced plastic responses being replaced by genetically determined responses is conveniently summarised by Ancel [4] as set out in Table 2.1. Using this framework, Phase 1 constitutes *'phenotypic accommodation'* [129] where the phenotype adapts to the environment through plasticity. Phase 2 is a change of distribution of alleles in the population towards the optimal phenotype. Phase 1 and 2 (and some would argue also sometimes but not necessarily including Phase 3) together represent the Baldwin Effect. Phase 3, where the population changes so that eventually the degree of plasticity reduces constitutes the *'genetic assimilation'*.

Phase	Phase Description
P1	The population wanders in phenotype space until a single individual has a norm of reaction that contains the optimal phenotype.
P2	Once the individual encounters the optimum, its lineage will quickly dominate the population.
P3	Selection will narrow the norms of reaction around the optimum.

TABLE 2.1: Adapted from the work of Ancel [4]. Three phases of the Baldwin Effect.

Waddington [120] provided the first experimental evidence of plastic responses changing evolutionary outcomes. His experiments showed that genetic assimilation could occur quickly: heat-shocked *Drosophila* larvae developed a proportion of *'cross-veinless'* winged adults and when these adults were subject to positive, artificial selection, cross-veinless flies appeared in the population without heat-shock within fourteen generations. The plastic response had been assimilated into the genetic code via changed selection pressures.

2.3.2 Genetic Assimilation through Canalisation

The argument for assimilation of the plastic response through canalisation is an important concept when considering the impact of plasticity on evolution. Waddington's original explanation, as discussed by Masel [72], is based on canalisation due to pleiotropic effects, where a 'canal' constrains development to produce a phenotypic trait - imagined by Waddington as a ball rolling down an inclined complex epigenetic landscape with peaks and valleys determined by genetic and environmental factors. An environmental input can push the ball from one canal to another. With this interpretation of genetic assimilation, selection for a trait-based on an environmental perturbation increases gene frequencies that deepen the 'canal' thereby requiring more extreme environmental inputs to change the expression of the phenotype (to move the ball into another canal). Over subsequent generations, the phenotypic expression becomes less environmentally sensitive to perturbation to a point where the environmental input has no effect. There are three key models built to explore the relationship between pleiotropic and epistatic effects and robustness of the phenotype against environmental and mutational perturbation: one by Masel [72], another by Siegal & Bergman [105] both briefly surveyed in Appendix A and one by Draghi & Whitlock [31] described in Section 2.7.2.

The canalisation theory of genetic assimilation has been subject to a different interpretations including a significant questioning of the common interpretation of Waddington's original explanation. A more nuanced description for a canalisation explanation for the Baldwin Effect, which has a bearing in our work, is proposed by Loison [68] who suggests that in his initial exploration of genetic assimilation, Waddington intended a more direct linkage between the environment and genes based on pre-canalisation of the plastic response. In this explanation, the change in phenotype in response to the environment can occur quickly and is a functionally appropriate phenotypic expression because the genotype/phenotype map has a pre-canalised path for development under those environmental conditions (for more detail see Loison [68]). This argument would seem to suggest a need for regularities in the environment so that pre-canalised responses have a high probability of being adaptive. It could be thought that Loison's pre-canalisation argument for the genetic assimilation is unlikely to hold for learnt behaviours discovered through an in-lifetime trial-and-error process, as these behavioural outcomes are unlikely to have been pre-canalised; however, physical traits that constrain behaviour could be seen as a form of internal canalisation, i.e. the repertoire of effective behaviours in a new environment will be constrained by physical traits. In addition, where learning discovers structural regularities in the fitness landscape, then incorporation of a representation of these regularities into the genome could lead to the pre-canalisation of appropriate behaviour.

This canalisation explanation for the Baldwin Effect is key to the hypotheses explored in this thesis and is discussed further in Chapters 5, 6 and 7.

2.3.3 Genetic Accommodation

The extent to which this adaptive plasticity reduces or increases as the plastically expressed traits get incorporated into the genetic code is also subject to debate. Therefore, the extent to which Phase 3 in Table 2.1 is fundamental to the Baldwin Effect is not certain. West-Eberhard [129] points out that Baldwin's original discussion of the inheritance of learnt behaviours also meant that phenotypic plasticity could increase under this effect. This potentially erroneous assumption of reduction of plasticity, West-Eberhard argues, may have led to a general underestimation of the importance of the Baldwin Effect and prefers Baldwin's original term *genetic accommodation*. She suggests that genetic accommodation, unlike genetic assimilation, may lead to a change of form, be genetically or environmentally induced, lead to an increase in plasticity, may increase environmental sensitivity and act on deleterious traits. As West-Eberhard [129] puts suggests, genetic accommodation does not necessarily entail a reduction in plasticity; the trait-mean value could change while the norm of reaction remains constant; under this scenario, the phenotype still reaches higher fitnesses.

The concepts of genetic assimilation, genetic accommodation and the Baldwin Effect continue to be often conflated in the literature. Crispo [22] attempts to disentangle the terms by providing a consistent definition of the phenomenon as reproduced in Table 2.2. However, the definitions of genetic accommodation offered by Crispo and West-Eberhard are so broad that it becomes difficult to differentiate between the categories and therefore do not offer much guidance as to what behaviours we might observe from a simulation of plasticity.

For the discussion in this thesis, the more general meanings for genetic assimilation and genetic accommodation proffered by Moczek [81] are used, where genetic assimilation requires a reduction in plasticity for all environments whereas genetic accommodation does not imply a reduction in plasticity. The term 'Baldwin Effect' is also used in its broadest interpretation where an individual's environmentally induced characteristics become heritable across a population through changes in selective pressures.

2.3.4 The Hinton & Nowlan Model

The first, and most well-known model purporting to show the Baldwin Effect is that constructed by Hinton & Nowlan [49] (for convenience we will call this the HN Model). In this model, the genotype is represented by twenty alleles controlling the connections *"in a neural net"* that are either decided and present (represented by a one), decided and

Term	Basis of trait inducing environmental response	Increase of or decrease in level of plasticity	Mean phenotypic value in the inducing environment
Baldwin Effect	Environmental	Neither or increase	Changes
Genetic Assimilation	Environmental	Decrease	Stays the same
Genetic Accommodation	Environmental or genetic	Neither or either	Stays the same or changes

TABLE 2.2: From Crispo [22]: Definitions for the Baldwin Effect, Genetic Assimilation and Genetic Accommodation.

absent (represented by a zero), or undecided (represented by a question mark). Each generation has two phases: an evolutionary phase where a non-overlapping population of children is generated by mating adults and a learning phase where undecided alleles are substituted with a one or zero. The operation of the model is briefly described below.

At the start of the experiment, a population of 1000 individuals is generated with randomly initialised genes, where on average, half the alleles are undecided, a quarter are decided and present (ones) and the remaining quarter are decided and absent (zeros). For a set number of pairings per generation, two parents are chosen probabilistically from this population, where the chance of being selected as a parent is proportional to the individual's fitness. A child is formed by crossing over the genes of the parents at two randomly chosen cross-over points. The population of the next generation is formed by the children generated. The number of children generated is equal to the number of individuals in the population and therefore the population size is constant. No mutation of alleles occurs.

The problem space requires reaching a target phenotype of twenty 'correct' connections, where all combinations other than a full set have correct connections has a fitness of zero and, a full set have a fitness dependent on the number of learning steps taken to achieve completely correct connections. The fitness landscape without learning is, therefore, a single fitness spike but smoothed by the action of learning that extends the basin of attraction.

Learning takes place as a random search on the undecided alleles for a pre-determined maximum number of learning trials. At each learning trial, one undecided gene is chosen at random and set to be either present (a one) or absent (a zero) with an equal probability. The fitness of the phenotype is tested after learning and retained if the fitness is higher. This learning trial is repeated until the target phenotype is found or the maximum number of learning trials is reached. The fitness of the individual is measured by the number of learning trials remaining once the target phenotype is found. Therefore, the quicker the solution is found through learning, the higher the individual's fitness. If the maximum number of trials is reached before the target phenotype is found, then the phenotype's fitness is deemed to be zero. The cost function in the fitness measure creates a selective pressure for 'undecided' alleles to be substituted by correct decided alleles. Once all individuals have completed their learning trials, a new evolutionary phase is started.

A re-implementation of the Hinton & Nowlan model using the same parameters reproduces the results observed in the paper, as shown in Figure 2.5.

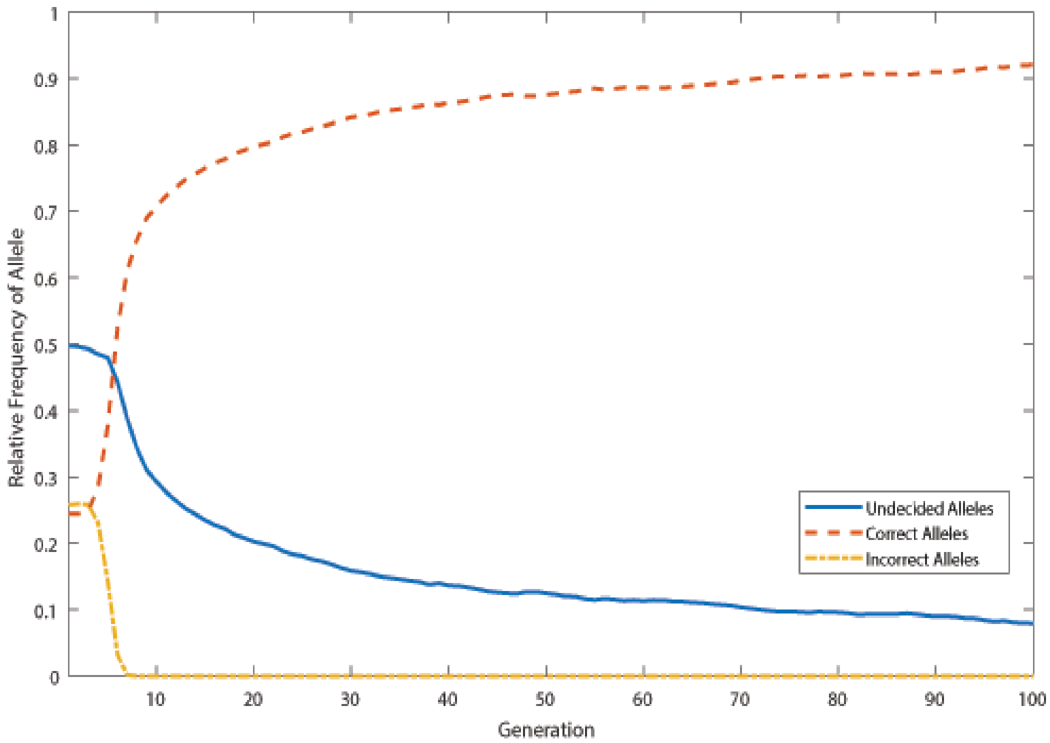


FIGURE 2.5: Reproduction of the results of the Hinton & Nolan [49] model for the Baldwin Effect. Plastic (undecided alleles) are reduced as the population finds the single global optima. In this simulation, there are 1000 individuals with 20 alleles. The simulation starts with on average 50% undecided alleles, 25% decided and incorrect, and 25% decided alleles.

Whilst being a straightforward demonstration of the action of genetic assimilation, the HN Model has three characteristics that inhibit it from being a good generalisation of the Baldwin Effect.

Firstly, the model conflates the representation of the genotype, the plasticity and the resultant phenotype into a single representation, where an allele can either be plastic or not. For example, Downing [29] citing Mitchell & Belew, points out that the genetic trait and the guessing of it by learning are interchangeable as there is no developmental process altering the mapping between genotype and phenotype.

Secondly, since an allele in the model can only either be plastic (represented by a question mark) or determined (represented by a one or zero), then the model structure together with the fitness function means that the plasticity of the phenotype must reduce, i.e. this can only be a model of genetic assimilation with no scope for there to be genetic accommodation. For a model to provide insight on the debate of whether plasticity is reduced as genetic code 'catches up' then the plasticity must be expressed separately to the *'breeding value'*.

Thirdly, the model assumes that the cost of in-life learning is less than the cost of the genetic change. As shown in Mills & Watson's [80] analysis of both the Hinton & Nolan and their model of the Baldwin Effect, from an engineering perspective, across the genetic change and in-life learning stages, the number of cycles required to find the solution is not reduced by learning (although in reality, in-life learning acts on a faster timescale than evolution). As pointed out by Fontananri & Meir [34], with the HN model, as the size of the 'genome' is increased, the time to reach the optima scales exponentially: so if the genome contained 100 alleles, then the solution would need on average 2^{99} cycles to find the peak (either learning cycles or evolution cycles) and therefore would not be consistently found within a reasonable amount of time. As described in much detail in Section 2.4 of this document, there have been a large number of other models of learning and evolution that challenge the conclusion that learning speeds the rate of evolution.

A more recent review of the HN model by Santos, Szathmáry & Fontanari [102] also challenges whether the HN model and Fontanari & Meir's [34] subsequent supporting analysis is a good representation of learning finding high-fitness configurations that evolution alone cannot. They have two main arguments to support this assertion. The first is that for a very large asexual population, the optimal genotypic configuration is discovered and becomes fixed in the population so rapidly that the impact of learning is negligible (they do this by considering the limit in an infinite population). The second is that the assumption that within a sexual population, good genetic combinations would be disrupted by sexual recombination and therefore will be lost without learning is flawed. They show that using different parameters the HN model without learning performs similarly to with learning (see Figure 3 in Santos, Szathmáry & Fontanari [102]).

2.3.5 Completeness of Genetic Assimilation

In the HN model, as can be seen in Figure 2.5, there is not a complete genetic assimilation of the learning with a few flexible alleles left in the population. This is due to there being little selective pressure to remove the last learnt *'genes'*. A more complex model showing partial genetic assimilation was developed by Dopazo et al. [28]. This study was a major advancement over the Hinton & Nowlan [49] model but still retained a similar representation for plasticity and the single-spiked fitness landscape. This model used the genome to encode the connections in a simple Ising perceptron[3]. Rather than ones and zeros and undecided alleles as in the HN model, this model used +1, -1 and flexible (represented by a question mark). The fitness function was defined as, given a random set of inputs, how many correct outputs were generated without the need for learning. The fitness being calculated as the number of input patterns where no learning was required to get the correct result as a fraction of the total number of example inputs. Like the Hinton & Nowlan model, the fitness was therefore dependent on the amount of learning required. The correct output was specified as having outputs of the same sign as a *'teacher perceptron'* (which was set to be all ones) for all sample inputs. The learning process took the genome and randomly assigned a value to those connections that were flexible (undecided in the language of Hinton & Nowlan). The output was compared against an example input pattern; if there is a difference in output and the target output, the signs of the flexible connections were flipped at random for a predetermined number of iterations. The method of evolution used fitness proportional selection, where fitness was assessed against a new set of inputs, followed by mutation without recombination. The model assumed a large population so that averaging of allele frequencies is valid. Unlike the HN model, where any zero allele causes zero fitness (because with any single zero the phenotype cannot be at the fitness spike and therefore has zero fitness), negative alleles (used instead of zeros) have a fitness contribution (as was the case with the Behera & Nanjundiah [8] models). Using this model, the authors found that in the later stages of evolution, the learning stops the evolutionary process from fully removing the plastic alleles; the more learning steps that are allowed, the greater this effect. The authors termed this a *'halting effect'* and this effect was explained by the few remaining plastic traits not creating enough selection pressure for complete assimilation due to the small fitness difference between ones and flexible genes. As the authors allude to, this dynamic is a function of the mechanism for calculating fitness: a negative one allele, has the same fitness value as an incorrect plastic allele and a positive one has the same fitness value as a correctly guessed plastic allele. The authors recognised that having a learning cost gives rise to an appreciable fitness difference between flexible and determined alleles and would therefore likely give rise to a different result.

What is not clear from the model described above, is what would happen in the case where there is a loss of phenotypic plasticity due to canalisation rather than an explicit cost of plasticity. Would full assimilation occur? This question is important for the models of learning in this thesis as real-world evidence would suggest that while there may be the genetic assimilation of learnt behaviours, it is unlikely that the learning element will disappear completely. For example, Hunt & Gray [50] suggest, based on their

[3]The Ising perceptron is a rudimentary form of neural network, with multiple inputs, one output with no hidden layers and a two-state input and outputs (-1 and +1).

previous work with juvenile and adult crows, that juveniles have an innate disposition for tool construction, but that adult crows can innovate and specialise in tool design in a way that is flexible and hierarchical. They conclude that *"Genetic assimilation does not eliminate learning, and with learning the potential for innovation is retained"*. Consequently, we might not expect to observe complete assimilation of learning in models of learning and evolution.

2.4 Behavioural Plasticity as an Accelerator of Evolutionary Change

As identified in Ghalambor et al. [38], there is much debate as to what extent adaptive plasticity and learning buffers or accelerates the impact of natural selection. The potential for plasticity to accelerate evolution is often referred to as the *Baldwin Expediting Effect*, as termed by Ancel [3]. The HN Model is widely considered to be a classical demonstration of both genetic assimilation and learning as an accelerator to evolution and, as discussed earlier, some research supports Hinton & Nowlan's conclusions. However, contrary to Hinton & Nowlan's claims, it is often argued that highly adaptive change, where the plastic response allows the phenotype to reach a fitness peak, is likely to buffer genetic change. This buffering is thought to be particularly pronounced where the plastic expression of the current allele frequencies within the population provides high fitness individuals, leaving no fitness differential for natural selection to act upon. This effect has been termed by Mayley [73] as a *'Hiding Effect'* and, as he points out, in the extreme case of all fitness being due to learning, there would be nothing for evolution to act upon.

Other analytical models and simulations have supported this challenge to the existence of the Baldwin Expediting Effect, with the debate as to whether phenotypic plasticity accelerates evolution primarily focussed on the shape of the fitness landscape. The key arguments and the models that support each side of the debate are reviewed in Sections 2.4.1 to 2.4.5.

2.4.1 Single-peaked Fitness Landscape

The first point of debate as to whether plasticity accelerates evolution considers a single-peaked landscape. Single-peaked landscapes can either be 'single spikes' like that seen in the HN model, where there is one single fitness peak with no fitness attributed to anything other than the ideal phenotype, or 'sloped', where fitness increases monotonically with the match to the ideal phenotype for an environment.

In support of the Hinton & Nowlan result for a single-spiked landscape, Fontaneria & Meir [34] undertook an analysis of expected gene frequencies using the HN model structure amended to utilise an asexual population subject to mutation rather than Hinton & Nowlan's sexual population without mutation. The key result was that, even with an asexual population, the Baldwin Effect is also observed if the population is given enough time to evolve. In addition, they found that learning insulated the phenotype from the potentially deleterious effects of a high mutation rate, effectively making the learning organism more robust. However, it would seem clear that this additional robustness is reliant on the fact that learning can directly substitute the allele value in the genome; this would be unlikely where there is a genotype-to-phenotype mapping that is more complex than the simple one-to-one map of the HN model and Fontanari & Meir analysis. In addition, as discussed in Section 2.3.4, later analysis by Santos, Szathmáry & Fontanari [102] suggested that some of Fontanari's original analysis was flawed for an asexual population because of inappropriate assumptions regarding linkage disequilibrium between alleles.

Lauren Ancel [4] used gradient analysis and simulations to challenge the theory that the Baldwin Effect accelerates the evolutionary process. In her model, a non-plastic population was represented as a uniform distribution of phenotypic values in a range and a plastic individual was defined by a range centred on the non-plastic phenotype's value. Mutations in the non-plastic population were made directly to the phenotypic value, whereas mutations to plastic individuals were made to the upper and lower bounds defining the range of plasticity. As per the HN model, the fitness landscape had a single fitness spike associated with the target phenotype with zero fitness elsewhere. The fitness function, for the plastic population, included a cost of plasticity based on the size of the norm of reaction (the larger the norm of reaction, the higher the cost). Using the phases in Table 2.1 as a reference for the different stages of the Baldwin Effect, Ancel's models showed that plasticity expedites the first part of evolution (Phase 1) as compared to a non-plastic population, where the population starts some distance from the optimal (either directly or within the norm of reaction). Once the plastic population encountered the optimum, convergence of the population was retarded as compared to a non-plastic population (Phase 2). The third step (selection narrowing the norms of reaction) is only required for a plastic population and therefore by definition, takes longer than a non-plastic population. One might argue that Phase 3 is not required for population survival and is not required where there is genetic accommodation rather than assimilation and therefore should not be included in the comparison. Ancel's framework for analysis is useful; however, the simulation, as with the HN model, suffers from a potentially unrealistic single-spiked fitness landscape.

Papaj's [90] simulations of learning also seemed to offer a contradictory conclusion to that of Hinton & Nolan. In his model, the behavioural response was a combination of an instinctual coefficient, a learning coefficient, and a defined number of learning experiences. In the experimental set-up, an asexual population was subject to evolution acting on each generation, with selection proportional to fitness and a set number of learning trials during an individual's lifetime. The fitness function was defined as a parabola based on the mean behavioural response, with a bounded fitness. Therefore, the mean behavioural response had an optimum value, and any response above or below that value led to a decrease in fitness. The fitness landscape was, therefore, single-peaked but of the sloped variety. Papaj's model showed that learning suppresses the evolution of innate responses because the individual can achieve maximum fitness through learning: the higher the learning rate, the greater the suppression of the speed of evolution. Papaj argues that there is biological evidence for this effect: for example, insects exhibit learning in feeding behaviours, even though, it being a basic behaviour, this would seem inefficient. Papaj goes on to argue that if 'learning' is heritable (which in this model is the inheritance of the rate of learning), then faster learning may evolve in parallel to the congenital response. In other simulations to test this hypothesis, Papaj claims that learning evolves at a rate that soon makes the evolution of the instinctual coefficient very slow. However, the graph to support this assertion (Panel A of Figure 6.6 in [90]) seems to show the instinctual and learning coefficients tracking together, and it could be the case that any retardation effect on the evolution of innate behaviours is sensitive to the rate at which mutation impacts the learning and instinctual coefficients. For an organism, it could be argued that the likelihood is that different alleles will control innate responses to those controlling the ability to learn and therefore comparing the mutation rates of the two is likely to be problematic (essentially comparing apples with oranges).

In addition to the criticisms above, Papaj's original model has three other key shortcomings. Firstly, as stated by the author, the behaviour, and consequently the fitness, is dependent on the learning equation where changes to learning value have an exponential effect and where changes to instinctual behaviours only have an additive value. Secondly, having the behaviour specified by an instinctual co-efficient at one locus and a learning coefficient at another locus leads to questioning of how to encode the interaction between these separate loci. While Papaj states that his method of combining plasticity and genetically determined values provides a learning curve commonly seen in animal behaviour studies, he admits that whether this has any real biological basis remains unknown. Thirdly, there is little insight into the potential costs of learning: while the author talks about the neural costs of learning, he does not consider the potential energy costs of exploratory behaviour associated with trial-and-error learning or the costs associated with sensing the environment. The absence of a cost of learning in this model would seem to explain why the evolution of the genome is retarded: there is no selective pressure for assimilation.

Another model, like Papaj's, that used a single-peaked, sloped fitness gradient was an early model constructed by Behera & Nanjundiah [8] where the fitness function was proportional to the number of ones and therefore rewards partial fitnesses. This retested some of Hinton & Nowlan's findings but this time using a sexual population evolving towards a target phenotype. Interestingly - and indicative of some of the confusion in the structure of the HN model and others based on it - the authors refer to the optimal phenotype as the target genotype. The fitness of a 'phenotype' after learning was evaluated as a weighted combination of the match of the phenotype to the target and a function of the number of learning trials, the fewer learning trials, the higher the fitness. Therefore, there was an explicit cost of plasticity. The maximum number of learning trials was tuned so that there was a reasonable chance that undecided alleles would become ones. At the start of evolution, the proportions of ones, zeros, and undecided alleles in the initial genotype were set so that the number of correct alleles (ones) was very small. Reproduction was via random selection from the set of highest-ranking individuals based on fitness and single-point crossover. The key result was that, with this particular fitness function, plasticity up to a certain level slows the rate of evolutionary change, as compared to without plasticity but increases the overall level of adaptation. It should be noted that, with this sloped fitness function, the genotype that fixes in the population is still some way away from the target, optimum phenotype. This model was a useful extension to the HN model, in that it showed the Baldwin effect might not be an accelerator of evolution where there is a smooth fitness landscape, but suffers from many of the same shortcomings: the primary one being the single representation for both the genotype and the phenotype.

Another extension to the HN model that considered a sloped fitness landscape was Morgan, Suchow & Griffiths [83] analysis of the nature of plasticity. In their model, they introduced a controllable mutual dependency between traits. This has the effect of creating a single-peaked, sloped fitness landscape (see Figure 2A. of Morgan et al's. paper [83]) which is either convex or concave in profile dependent on the degree of dependency between alleles. In this model, the original HN experimental set-up is an extreme case where there is total mutual dependence between alleles, i.e. all traits need to be correct for there to be a greater than zero fitness. Morgan et al's key result is the higher the mutual dependence between traits, the more quickly the traits will become fixed (non-plastic) in the population. They also found that where there is a chain dependency between traits - each trait is dependent on one proceeding trait so that the second trait is dependent on the first trait and the last trait is dependent on all other traits - the likelihood of fixation of a trait is a monotonically decreasing function of the number of traits that depend on it. Interestingly, this model did not have an explicit cost of plasticity - the only cost was the potential mismatch between the phenotype and the environment - this is in contrast to every other Hinton & Nowlan based model that explicitly included a cost of plasticity. However, the convex and concave fitness landscape profiles created by the trait dependence creates a smooth fitness gradient towards an optimum and therefore the genotype will move towards this optima.

As well as trait dependencies, unlike the HN model, this model also allowed for plasticity to be inherited through social learning but since this is ignored in this thesis, this result is not discussed here.

2.4.2 Multi-peaked Fitness Landscape

Multi-peaked fitness landscapes are thought to be more representative of the fitness challenges to real organisms than single-peaked landscapes. For example, the work of Niklas [86] on modelling the evolution of plants during the Devonian period suggests that the more facets of environmental demand on the organism (*'the more biological tasks to be simultaneously performed'*), the more local optima will appear in the fitness landscape. Interestingly, this work also suggests that the average height of the fitness peaks will reduce as phenotypic complexity increases. This aligns with Frank's analysis [36] that suggests that plasticity in multiple dimensions is likely to have lower fitness peaks. This is discussed further in Section 2.7.1. Therefore, there has been significant interest in examining the effect of plasticity on rates of evolution in these landscapes.

Borenstein, Meilijson & Rupin [12] undertook a mathematical analysis of the potential interaction between phenotypic plasticity (specifically learning) and evolution. This analysis introduced the concept of *'drawdown'*, the largest descent in a multi-peaked landscape. The simulations in this work showed a strong correlation between the depth of the largest draw-down and the time it takes to find the global optimum. Together with random walk analysis, this correlation allowed a bound to be calculated for the predicted rate of evolution for any landscape. In this analysis, the alterations due to learning were made to the phenotype with a one-to-one mapping between genotype and phenotype. Learning was modelled as a deterministic hill-climbing algorithm where the innate phenotype (without learning) is moved towards the nearest local optimum. The analysis showed that in the scenario where learning proceeds until a local optimum is encountered (termed *'ideal learning'*), the effective fitness landscape becomes stepped with the size of the step equal to the difference between the heights of the peaks. This stepping reduces the size of the drawdown, and therefore, the authors point out, accelerates the rate of evolution in a multi-peaked landscape. This ability to reduce the size of the drawdown is particularly interesting in the scenario where the size of the drawdown would make moving to the global optimum highly unlikely because the population would effectively be stuck at a local optimum. While Borenstein et al. do not explicitly consider this scenario, it is discussed further in Section 2.5. The paper also showed, aligned to Papaj [90] and Bereha et al.'s [8] results, that in a single-peaked landscape, where there is no fitness valley, no acceleration is observed. In addition, in these landscapes, the learning flattens the fitness landscape, and therefore population convergence is slower. Learning that was limited so as not to achieve the local optimum (termed *'partial learning'*) was found to have an acceleration effect but to a lesser extent than *'ideal learning'*. Borenstein et al. also researched a stochastic learning algorithm based on the technique of simulated annealing. They found that this type of learning also accelerated the rate of evolution towards the global optimum as compared to non-plastic evolution but that the resultant flattening of the fitness landscape improves population convergence as compared to deterministic learning. The authors admit that there are key shortcomings to their analysis, the primary ones being the lack of a cost of learning and the simplistic one-to-one map between genotype and phenotype. This model is discussed in detail further in relation to a cost of learning in Section 3.2.2.

2.4.3 Varying Fitness Landscape

In common with a multi-peaked fitness landscape, a fitness landscape that has temporal variability presents challenges to plastic adaption to the environment. One investigation that used varying fitness landscapes was conducted by Lande's [63] analysis of phenotypic plasticity. This simulation went beyond Ancel's concept of representing the norm of reaction as an upper and lower bound of phenotypic trait values with the now popular representation of a reaction to environmental input as an elevation and slope, where the elevation is the mean phenotypic value - the sum of the average effect of alleles (the *'breeding value'*) - and the slope plots the degree of plasticity in each environment. His mathematical treatment compared plasticity between one and two environments where for each genotype, its phenotype is a function of the environment in which it evolved. In a static environment, Lande presents the accepted thinking as that a small change in the environment would mostly cause a phenotypic reaction by changing the slope, whereas large changes in the environment would cause shifts in elevation. His analysis examined the movement in the slope and elevation following a large-scale change in environmental conditions, primarily by comparing variable plasticity (plasticity is varied by selection), constant partially adaptive plasticity (plasticity not subject to selection) and no plasticity. The results show two distinct phases of reaction to environmental change following what was at first, as one would expect, a large drop in fitness as the phenotype no longer matched the optimum phenotype for the new environment. During Phase 1, selection causes a big jump in phenotypic plasticity and subsequent fitness, as compared to constant or no plasticity, with a slower rise in genetic contribution as compared to static or no plasticity. The previously uncorrelated slope and elevation of the norm of reaction become strongly correlated during this phase. During Phase 2, the no plasticity and constant plasticity scenarios catch up with selectable plasticity in terms of phenotypic fitness between generations and plasticity sinks back down to the previous canalised level after that. With variable plasticity, the genetic contribution to the phenotype levels out to the same level as constant plasticity but at a lower level than no plasticity. During this phase, the selective direction for plasticity, and the breeding value are opposite, and therefore the plastic expression becomes genetically assimilated. Lande suggests that over millions of years of evolution, species are subjected to multiple large-scale shifts in the environment, and therefore plasticity is critical to long-term survival. Whilst providing insight into large environmental shifts, this quantitative genetic approach is mostly constrained to considering a single trait and therefore may not be realistic when considering behavioural plasticity (see Section 2.7.1).

A similar analysis by Saito et al. [101] tested their model of plasticity that used a Gaussian distribution to represent variance in population plasticity (see Section 3.2.3 for more detail) in a periodic fitness landscape; the results of both the mathematical treatment and individual-based simulations showed that the average fitness was reduced as a non-linear function of the degree of environmental fluctuation. In addition, the analysis and simulation showed that the higher the phenotypic fluctuation, the more likely the genotype could 'jump' from one fitness peak to another through smoothing of the effective fitness landscape.

2.4.4 Environmental Complexity

Fitness landscapes are usually assessed either in terms of their ruggedness, or the speed at which the environment changes. These aspects are an important consideration for any model of learning and evolution, as discussed below, this can critically affect learning performance.

Dridi & Lehman [32] used a mathematical simulation to suggest that not only environmental change over a lifetime but also environmental complexity is an important part of the balance between innate behaviours and learning. In their model, they considered individuals in a population that experienced a parameterised subset of possible stimuli in an environment and a controlled switching rate of the environment. An individual in a population had a memory of a fixed size that could be used to store a variable number of innate responses or learnt responses. The trade-off between innate responses and learnt responses was an evolved characteristic with a higher payoff for getting a correct innate response than a correct learnt response, which in turn has a higher payoff than correct random responses. By considering different environmental complexities and switching rates, their analysis suggested that environmental complexity will favour the evolution of learning. Dridi & Lehmann characterise the case for learning being favoured in heterogeneous environments as being built on the fact that if the next generation encounters a truly novel environment, then the behavioural responses encoded in the genome (i.e. the innate responses) will not be appropriate to that environment. Therefore, individuals driven by innate behaviours will not possess the appropriate behavioural responses and will die out; but individuals that have evolved mechanisms that support learning will be favoured. This would suggest a lineage selection [97] account for the evolution of learning.

While providing an interesting result, this model has a potential shortcoming. The key assumption of the Dridi & Lehman model was that the innate and learnt behaviours compete for control of the phenotype's response (a behaviour is either innate or learnt). However, this is contrary to much of the debate highlighted in Section 2.1.2, where learnt and innate behaviours act in concert to produce a blended response. This competition for control is a similar issue to the HN model having learning and genetics competing for control of the phenotype. For the model of learning and evolution presented in this thesis, innate behaviours and learning should be able to influence the resultant behaviour but not directly compete for complete control of the phenotype's response.

As discussed later in Section 3.3.2, Mayley [74] used Kauffman's NK landscapes [55] to control environmental complexity. However, these landscapes possess little structural regularity with the fitness values of local and global optima being difficult to determine. Therefore these landscapes constrain explanatory power and may not be a good option when considering a correlations model of plasticity.

2.4.5 Indeterminate Fitness Landscape

Many simulations of the interaction between learning and evolution have been developed by researchers in the Artificial Life (ALife) community. While models created by quantitative geneticists and evolutionary biologists tend to consider the interactions between a small number of traits and simplistic forms of plasticity, ALife simulations tend to consider multiple traits and have biologically inspired learning mechanisms such as Artificial Neural Nets (ANNs). These ALife simulations are also very different from traditional analysis and simulations in that, in common with empirical work, they are open-ended; with no predetermined fitness landscape or 'optimal' end-state for evolution to reach. They, therefore, as pointed out by Menczer & Belew [76] citing Gould and Lewontin, do not treat evolution as an optimization task but more as an ongoing process where populations can adapt to the environment. This approach chimes with Diogo's [27] view that evolution should not be considered to be an optimizing task that has an optimal outcome. For ALife models, the success of the population is measured in different ways to that of typical adaptionist models of evolution, i.e. those that have defined fitness optima, because there is no target phenotype and therefore no optimum fitness. Measurement methods include the frequency of a type of individual in a population (e.g. plastic versus non-plastic) or how long the population can survive given a set distribution of 'resources'.

Despite the complexities, these models do claim to show interesting interactions between learning and evolution. Of the seven ALife inspired models that examined the interaction of learning and evolution uncovered as part of a literature search, two claim that learning does not improve evolution [29, 93], one suggests that learning improves evolution for certain types of learning [76] and four claim that learning improves evolution [1,37,65,87]. However, the lack of known local or global fitness optima and general indeterminacy of the fitness landscape makes it highly problematic to assess whether phenomena such as the Baldwin Optimizing Effect discussed in the next section, are being observed.

2.4.6 Experimental Evidence For Accelerated Evolution

The simulation work discussed in this chapter suggests that whether plasticity in general and learning in particular, accelerates evolution is dependent on the shape of the fitness landscape. In common with the Artificial Life simulations discussed in Section 2.4.5, the exact shape of the fitness landscape is also not known for empirical experiments testing phenotypic plastic response. However, field experiments do provide insight into whether plasticity accelerates or retards evolution. One such example is Ghalambour's [40] fieldwork on guppies (*Poecilia reticulata*) with cichlid predators alongside the analysis of gene transcription factors. This work showed that generally plastic gene expression is in the opposite direction to adaptive evolution, i.e. adaptive plasticity constrains evolution whereas non-adaptive plasticity accelerates adaptive change. However, interestingly, this experiment chose to move the guppies from an environment with predators to one without predators. If the experiment had been the other way around, moving from a non-predatory to a predatory environment, the guppies would have been unlikely to survive with a non-adaptive plastic response.

2.5 The Baldwin Optimizing Effect

In the introduction to their paper, Zollman & Smead [133] identify three components of the Baldwin Effect; the Simpson-Baldwin effect, the Baldwin Expediting Effect (as named by Ancel [4]) and the Baldwin Optimizing Effect. In the authors' conceptual framework, the Simpson-Baldwin effect is simply genetic assimilation and the consequent reduction is plasticity, and the Baldwin Expediting Effect [3] is the potential for plasticity to accelerate evolution as discussed extensively previously. However, the Baldwin Optimization Effect, a concept very relevant to this research, is not so widely discussed and relates to phenotypic plasticity changing evolutionary trajectories with the potential to find fitness optima that would not be found without plasticity.

An example of an investigation of the Baldwin Optimizing Effect, cited by Zollman & Smead [133], is that of Mills & Watson [80]. In their mathematical analysis and accompanying computational simulation, the authors looked at how the Baldwin Effect can facilitate a learning population to avoid getting stuck on a local fitness optimum and therefore *'cross fitness valleys'*. In this instance, they deployed a two-peaked fitness landscape, where one peak is a local optimum, and one is the global optimum. In their model, unlike the Hinton & Nolan [49] model there was no inbuilt reduction of plasticity where a trait is either plastic or not; instead, the phenotype was a local adaptation of the genotype values, where the learned phenotype is the best adaptation found. Their results showed that with a relatively high number of lifetime learning trials, the learning could smooth the fitness landscape so that there is a constant direction to the fitness gradient to the global optimum. The authors acknowledge, in common with the Hinton & Nolan model, that there is no engineering benefit to learning in terms of reduction of the number of fitness evaluations required to reach the global optima, but importantly the population reaches a fitness peak that would not be attained without plasticity.

Zollmann & Smead [133] set out to further examine which of the three facets of the Baldwin Effect they identified would be observed in their model of the evolution of plasticity. Their investigation simulated the learning of language and built on their previous work on plasticity in cooperation games. In this model, the population is comprised of pure-play strategies and reinforcement learners. In their experimental set-up, based on the Lewis signalling game, there were two types of agent: a signaller and a responder. The signaller transmitted information about the environment and the responder received the signal but not any direct information about the environment. The responder could choose from a set of behaviours only one of which is appropriate to the environment. If the responder chose the right behaviour, both the signaller and the responder received a reward. Otherwise, both parties received no reward. While in this set-up there is an optimum where the signaller sends the right signal and the responder behaves appropriately, there are also sub-optima where the signaller chooses to send the same signal, all or some of the time regardless of environmental state and therefore the responder's best strategy is to behave according to previously correct actions. Their key finding was that the Baldwin Effect was observed with plastic learners being selected for at the start of evolution (i.e. Ancel's Phase 2) and then selected against as pure-play strategies dominate. This evolutionary profile was because there was a greater payoff for a fixed signalling strategy when interacting with a learner (akin to a narrowing of norms in Ancel's Phase 3). However, the authors also found that the initial presence of learners enabled the optimizing effect where the population converges on a global optimum.

Both the Mills & Watson and Zollman & Smead models show plasticity can alter evolutionary trajectories. While presented here as discrete concepts, the Baldwin Optimizing Effect and the Baldwin Expediting Effect are inextricably connected: the size of the fitness valley affects the speed of evolution due to the relative probability that individuals within a population can escape that valley. For a given constant mutation size, the larger the valley, the less the probability of jumping from one peak to the next, to a point where the phenotype is stuck at a local optimum. It is when the population would normally be stuck at a local fitness optimum, any mechanism to move the phenotype becomes an optimizing effect rather than an expediting effect.

2.6 Behaviour Enabling Population Robustness & Persistence

In the traditional account of the Baldwin Effect, the plasticity provides higher fitness phenotypes (phenotypic accommodation) and then selection can act of existing or new variation to move the genotype closer to a local optimum [22]. The extreme case is of phenotypic plasticity allowing the population to survive rapid changes to the fitness landscape occurring either through major perturbations to the current environment or to facilitate the invasion of a new environment. This is true for learnt behaviours as well as innate behaviours. Behaviours have enabled species to survive in novel environments as adaptive behaviours within a lifetime mitigate the effects of selection by the environment.

When examining phenotypic plasticity in the context of sudden changes to the environment, Frank [36] introduced the concept of *'fitness truncation point'*, the point in a fitness landscape where, if a plastic population were entirely below a certain level of fitness in that environment, it would become extinct. This relates the truncation point to the population surviving during the first step in Ancel's three-step description of the Baldwin Effect (as shown in Table 2.1). Frank argues that a broad norm of reaction should be favoured, as this makes the population more likely to survive a major shift in the environment because it is more likely that some of the population will be beyond the fitness truncation point. This would apply to both single and multi-peaked fitness landscapes.

The model that Frank [36] deployed to illustrate this point was as follows: the reaction norm was represented as a probability of phenotype being x given a reaction norm centred at $\bar{x}$ (with a normal distribution with mean $\bar{x}$ and variance γ^2). The fitness of phenotype x was normally distributed (with a mean of 0 and variance σ^2). Like other model's of plasticity, Frank found that a norm of reaction smooths the fitness landscape and also showed that broad reaction norms enabled movement between fitness peaks so that an optimum fitness can be found and this is true for either discrete or continuous values. He also showed that in extreme or novel environments, only phenotypes beyond the truncation point survive and therefore broad norms of reaction are favoured under these conditions.

However, the survival of the population is not solely dependent on plasticity. The fitness truncation point can simply be viewed as an extreme case of the fitness differential that plasticity can provide. If a plastic population can gain additional fitness through adaptive plasticity, then natural selection will, over time, ensure that phenotype will become more prevalent in the population as per Phase 2 in Table 2.1. If that fitness differential is needed to survive, then the phenotype will dominate the population very quickly. The fitness truncation point is also a concept that applies to all population genetic diversity: as Frank [36] recognises, having a diverse population of genotypes also produces a wide range of phenotypes, therefore, making it more likely for a proportion of the population to survive a major environmental perturbation.

Brown [14] also argues that learning can alter evolutionary trajectories through the maintenance of genetic diversity without recourse to arguments for the incorporation of the learning into the genotype or the exo-genetic inheritance made possible by social learning. She hypothesizes that the fitness shielding effect of learning (a la Mayley's hiding effect discussed earlier) helps maintain population genetic diversity leading to improved lineage robustness in the face of further changes to the environment. Her case is supported by contrasting the evolutionary response to toxic cane toad *(Rhinella marina)* invasion by native species in Australia. The red-bellied black snake *(Pseudechis porphyriacus)* population has been observed to suffer extensive mortality when invaded by cane toads, but due to this strong selection pressure, the population quickly recovers with snakes that have a latent resistance to the cane toad toxin dominating. In contrast, the mouse-like planigale *(Planigale maculata)* learns to avoid predating the cane-toad after non-fatal encounters through the conditioned response of taste aversion. The planigale is therefore not subject to the same strong selective pressures for toxin resistance. Brown, therefore, suggests that the result of the learning is to have more genetic diversity. This would seem to be the opposite of the case for the Baldwin Effect.

2.7 Learning Acting on Multiple Traits

2.7.1 Multi-dimensional Phenotypic Plasticity and Behaviour

As discussed previously, many models of plasticity are single or dual loci. However, as Westneat [131] expounds, just considering one plastic trait can be a misleading simplification. Building significantly on the work of Chevin & Lande [17], Westneat introduces the concept of multi-dimensional phenotypic plasticity (MDPP) which considers how multiple plastic traits can affect evolutionary outcomes. MDPP recognises that environmental effects on traits have many complex facets: multiple environmental factors can vary continuously or discontinuously and traits can respond to these environmental factors via continuous or discontinuous reaction norms to the cue. Besides, the plasticity of one trait can impact plasticity in another (multi-variate plasticity) forming *"cross-trait correlations in plasticity"* [131].

Environmental effects on traits can be additive (and subtractive) or non-additive. Adapted from Westneat [131], a simple illustration of additive and non-additive effects of MDPP is shown in Figure 2.6.

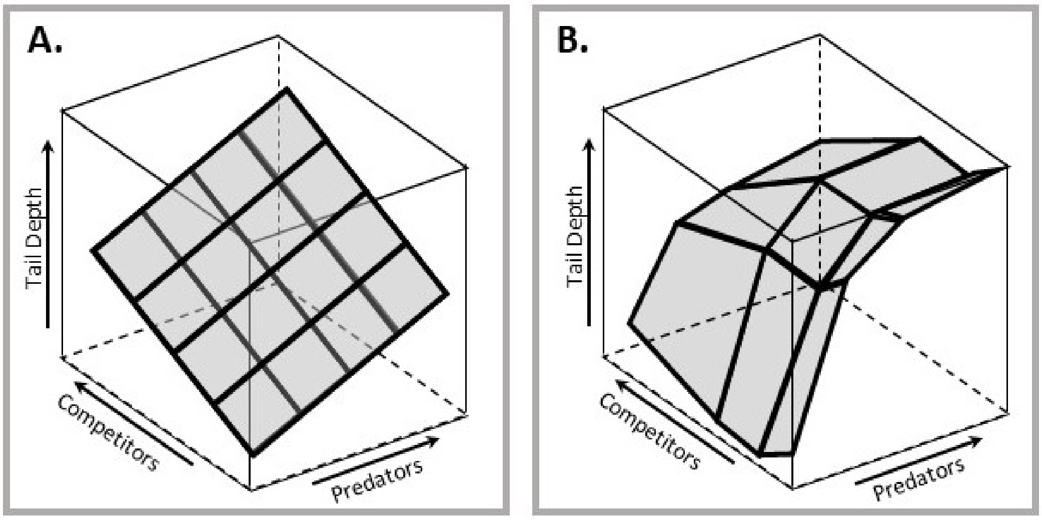


FIGURE 2.6: Adapted from Westneat [131]. **A.** An additive norm of reaction. **B.** A non-additive norm of reaction.

Westneat suggests that MDPP will have profound effects on selection. Individual traits may confer differing benefits under different environmental conditions (this includes trade-offs, e.g. searching for food vs exposure to predators). Therefore, there is a potential advantage to multiple cues changing plasticity as it can provide the ability to tune plastic response to the environment (including trading-off costs versus benefits of plasticity). For example, slope dependent costs (cost increases as the magnitude of plasticity increases) may be limited if a second cue correctly limits plasticity (and therefore the cost associated with that plasticity). In addition, MDPP may lead to improved precision in sensing the environment, allowing better plastic response to the environment thereby providing higher fitness. Costs of precision may be critical to shaping the reaction norm (for example, sequential effects may dramatically increase imprecision), therefore selection can be strong for non-additive MDPP.

MDPP could help survivability in changing environments: if one cue fails to trigger the right plastic response then another cue might still provide the correct plastic response. Westneat cites the example of light pollution from land affecting turtle hatchlings' ability to find their way to the ocean. Westneat hypothesises that if a second cue were also available, for example, the sound of the ocean, then this second cue may improve the survival rate of those using that cue. Over time, there would be a shift from one cue to the other (selection would favour sound over vision for orientation). Westneat also argues that the same argument applies to new environments, multiple cues lead to robustness in response by generalising the response over multiple stimuli. Interestingly, Westneat chooses to use examples where the plasticity in question is behavioural suggesting that multiple dimensions to behavioural plasticity.

Similarly, in learning Frank [36] suggests that multi-trait plasticity is likely to have a greater dilution effect on the fitness of plastic individuals as compared to the fitness of non-plastic individuals. He used the example where the reaction-norm varied across multiple traits and claimed that the smoothing effect becomes more pronounced reducing peak fitnesses as compared to the peak fitnesses with single-dimensional plasticity or as compared to the original fitness landscape.

The arguments presented above would suggest that a model where plasticity is expressed across multiple traits would lead to significant reductions in fitness thereby negating any optimizing effect. However, common to both the Saito et al. [101] and Frank [36] models is the way that the Gaussian distribution of plasticity is applied to the population of genotypes. Both models did not account for the scenario where selection on the fitness gain enabled by plasticity would lead to that genotype dominating the population (as per Phase 2 in Ancel's description of the Baldwin Effect, described in Section 2.3.1) thereby tightening the distribution of plasticity around peak fitness increasing overall fitness.

The above suggests that plasticity across multiple traits (including multiple behavioural cues) has the potential to have a significantly different evolutionary dynamic than the mostly single or dual loci models reviewed in Sections 2.4.1 and 2.4.2.

2.7.2 Plasticity in Complex Genotype to Phenotype Maps

The single and multi-peaked models of evolution with plasticity reviewed in the previous sections represent a plastic contribution to the fitness value as either a norm of reaction that is additive to or a multiplier of, a phenotypic trait value. The plasticity is defined explicitly by the genotype and most of the adaptionist models of evolution, reviewed earlier, are single locus or dual loci models. However, it is broadly recognised that real genotype-to-phenotype maps have multiple loci and more complex interactions between genotype and phenotype, primarily due to epistasis and pleiotropy. Real-world experiments show that traits can be a product of tens of genes, for example, even simple traits such as *Drosophila* eye colour are the product of over 70 genes alongside complex epigenetic interactions [129]. In this context, consideration of the impact of the environment on phenotypic plasticity requires a richer representation.

One model that had a more complex mapping between genotype and phenotype than the HN model was Behera & Nanjundiah's [10] second major simulation which adapted their previous models to introduce the concept of environmentally sensitive regulatory genes. This model enabled the concept of canalisation of phenotypes to be explored.

They used Waddington's [119] experiments on the heat-shock of *Drosophila* as the inspiration for the experimental set-up. This model embodied two environments; environment A corresponded to the common-garden environment whereas environment B corresponded to the heat-shocked environment. Regulatory genes were always determined, whereas structural genes could be plastic (i.e. structural genes had the values 0 or 1 or X for undecided). The frequency of ones in the regulatory genes controlled the probability that an undecided allele in the structural gene was to be assigned a one, where this probability increased proportional to the number of ones in the regulatory genes (see Behera & Nanjundiah [9] for more detail of this mechanism). In environment A, the probability that regulatory genes were ones was significantly smaller than in environment B and this skewing, therefore, indirectly affected the degree of plasticity of the phenotype in that environment. To verify that genetic assimilation did indeed take place, the offspring of each generation was divided into two groups, the test group (T) was exposed to the stressed B environment, and the fittest phenotypes from the other group that were siblings of the best performers in the T group were selected to form the next generation. The main result was that, like Waddington's empirical work, genetic assimilation occurs when there was selection for the alternative environmentally induced phenotype with a subsequent loss of plasticity in a small number of generations. The authors admit that their choice of parameters, mainly the extent to which the environment influences the number of ones in the regulatory loci, were tuned to give a result in broadly the same number of generations in which the genetic assimilation was observed in Waddington's experiments. Although focussed on developmental plasticity rather than learning, this model offers some advancements as compared to the original HN model: the plasticity was environmentally sensitive, and the separation of the structural and regulatory genes enabled a more complex mapping between genotype and phenotype. However, while this model aimed to replicate Waddington's laboratory experiment, the one-to-one mapping between regulatory gene and structural genes means that there is no epistasis or pleiotropy and therefore is unlikely to represent the genetic structure encoding *Drosophila* wings adequately.

A perhaps more biologically realistic model, with a complex mapping interaction between genes, was constructed by Draghi & Whitlock [30]. That work tested the hypothesis that phenotypic traits will express greater mutation variance, genetic variance and evolvability along trait dimensions that have more plasticity. To test this assertion, they constructed a model that implemented the concept of a network of genes that had three layers. The first layer was subject to environmental cues (sensor genes), the second layer took inputs from the sensor genes and also regulated each other through transcription factors, and these, in turn, regulated a third layer of genes that controlled the expression of a phenotypic trait. In this model, there was both a binary encoding as to whether two genes are connected and also a representation of the connection strength. Both types of connection were subject to mutational variance, with connection strengths retained in the model even if connections were switched off through the binary encoding. The 'development' of the traits was achieved through a twenty time-step process that iteratively applied the weight of inputs of actively connected genes. Trait optima were defined by a fixed value or in the alternative, heterogeneous scenario, a value that varied with each generation where the trait optimum was the same for both traits. In this model, therefore, the plasticity was represented by the different phenotypic values expressed when there was different environmental input to the environmental sensing genes. Fitness was calculated as the Euclidean scaled distance between the actual and optimum trait values. The parents of the next generation were probabilistically selected from a population-based on the fitness with the offspring generated using recombination of the linkage and connection strength pairs from each parent. The results showed that for a heterogeneous environment, the network of genes evolved greater adaptive plasticity than in a static environment. The authors also found that phenotypic plasticity led to greater mutational variance and standing genetic variance than when the environmental cue could not be sensed. A major strength of this model was that there was no need for direct encoding of plasticity, as there is with most other models previously discussed. Environmental cues were allowed to influence the phenotypic traits moderated through the regulatory network, and therefore, plasticity was expressed as a consequence of the dynamics of the system, i.e. individuals evolved that were close to the trait optima in multiple environmental conditions cued from the fitness landscape. However, this model was still limited by having only one continuous variable as the environmental input and only two phenotypic traits.

The models discussed in this section all are significant improvements on the models that have simple structures akin to that found in the Hinton & Nolan model and avoid the nuances of biological realism clouding the clarity of results. However, although the mapping between genotypes and phenotypes is complex, they are still limited to considering a small number of traits, which in itself is likely to be biologically unrealistic.

2.7.3 Correlations in the Genotype

In a review of empirical studies across 22 species, Roff and Fairbairn [100] suggest that the selection gradient on the correlation between two traits is of the same sign as the correlation between the traits and also there is a positive correlation between the strength of the correlation selection gradient and the correlation of the traits. From this, it might be deduced that over multiple generations correlations between traits become increasingly strong in either a positive or negative direction.

One physical manifestation causing correlations in the genotype is that of Gene Regulation Networks (GRNs), where genes can up and down-regulate each other during the process of phenotypic development. Many empirical experiments and simulations of the evolution of GRNs show that these networks help maintain robustness against deleterious mutations and increase the effectiveness of beneficial mutations, thereby increasing the evolvability of evolution [23]. Furthermore, there is evidence for genetic accommodation in correlations: experimental evidence conducted by Waddington, as cited by Crispo [22], that involved exposing *Drosophila* to ether suggests genetic accommodation can also act on regulatory networks in the same way that it works directly on traits. West-Eberhard [129] also emphasises the importance of polygenic traits when thinking about genetic accommodation: she cites examples where even simple traits such as *Drosophila* eye colour are the product of over 70 genes alongside complex epigenetic interactions.

However, the examples above suggest a physical linkage between genetic correlations. Would this necessarily hold for learning where a learnt behaviour impacts multiple, potentially disparate, traits?

There are other forms of genetic correlation that are not necessarily caused by the chemical linkage between the actions of genes. For example, if selection favours front legs and the back legs of a quadruped to be the same length then selection forces rewarding long back-legs, will also reward long front-legs. As Roff and Fairbairn [100] point out, the correlation between bivariate traits in the various empirical studies that they reviewed may be due to either pleiotropy or other causes of linkage disequilibrium.

For our simulations, we assume that correlations between traits evolve based on selective forces mediated by learning which may be due to a combination of pleiotropic or epistatic effects, or alternative causes of linkage disequilibrium; this builds on the work of West-Eberhard [129] described in more detail in Section 2.7.4.

2.7.4 West-Eberhard's Correlation Effects of Learning

West-Eberhard [129] suggests that learning acts to form correlations between physical traits. Using the example of beak morphology and food selection learning in birds, she supposes that there is a potential for a positive feedback loop (*'circular reinforcement'* [130], leading to correlations between alleles. The learning acts as a fitness screening mechanism mediating the selection forces on the genotype.

The logic of her hypothesis, with West-Eberhard's examples of dietary choices in birds, is represented in the diagram shown in Figure 2.7.

While learning cannot be directly transmitted between generations, West-Eberhard [129] submits that it is the recurrent trait complex appearing in each generation, for example, the beak-behaviour-dietary-gut complex, that is the unit of selection. This notion is supported by Morgan et al. [83] who use the ability of Archer fish to squirt water to catch flying insects as an example of mutual reinforcement between the learnt squirting behaviour and eyes that can accurately see objects above the water. Similarly, Snell-Rood

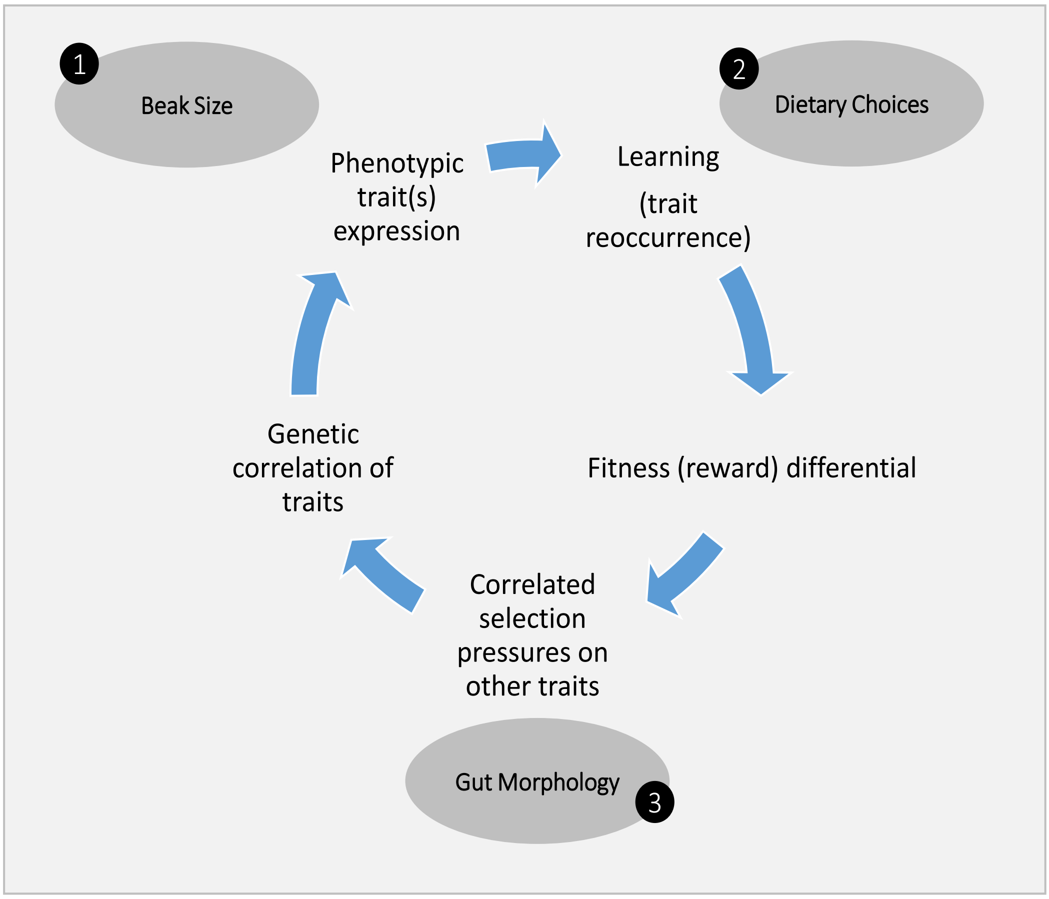


FIGURE 2.7: A diagram interpreting West-Eberhard's explanation for the correlation effects of learning. Learnt behaviours create a fitness differential that mediates selective forces on a range of physical traits leading to a correlation between genetic traits.

& Steck [110] identify that physical traits that support behavioural traits can form feedback mechanisms: they cite the example of attention to UV light detection which can, through simplification of cues, make what appears to be a complex search environment into a simple one, making food sources easier to identify. In this case, there is selective pressure for the acuity of UV light detection in the context of foraging behaviour. Building on this concept, it can be hypothesised that the changes in morphology induced by previous generation's learning will also make the recurrence of that behaviour more likely. For example, if learning to jump to feed on fruit is more beneficial with larger hind-leg muscles and an improved sense of balance, then the evolution of those larger muscles and balance improvements is more likely to produce a reward when the next generation learns to jump again through trial-and-error. Moreover, in this example, the trial-and-error learning is likely to be rewarded earlier, i.e. jumping becomes easier.

Corning [21] echoes West-Eberhard's ideas by proposing that behaviour, and presumably learning, is beyond phenotypic plasticity because it has intentionality and direction, what he terms *'Teleonomic Selection'*. Using the previously discounted Lamarckian example of the continued stretching of a giraffe's neck to reach the tops of Acacia trees causing heritable changes, he postulates that a population of giraffes' newly acquired habit of reaching for acacia leaves causes a fitness screening effect that will provide an additional fitness benefit to giraffes with a longer neck.

2.7.5 Learning Correlating Innate Behaviours

In this thesis, I hypothesize that West-Eberhard's correlation theory of learning described in the previous section can apply to innate behaviours as well as physical traits. Learnt behaviours can apply selective pressure across several innate behavioural traits, thereby correlating them. If the combinations of correlated innate behaviours make the learning easier, then there will be selective pressure for the learning to be assimilated (as combinations of innate behaviours). Once innate, the recurrence of the behaviour between generations becomes highly probable, and therefore there will be a more consistent direction of selection. Therefore, genetic assimilation of learnt behaviours has scope to reinforce the virtuous circle shown in Figure 2.7.

An important characteristic of this emerging model for learning and evolution is that the physical traits correlated by the learning are not necessarily the same alleles controlling the learning and behaviour. Using the bird example above, the feeding behaviours change the selection pressure on many traits and these traits are not directly related to the behaviour, i.e. the genes that control the phenotypic traits are not the same traits directly controlling the learning and, as discussed in Section 2.1.7, whilst learning is genetically constrained it is not genetically determined. This point is also true for learning and innate behaviours: the genes controlling the mechanisms of learning are unlikely to be the same as those controlling innate behaviours - learnt and innate behaviours have distinct, yet overlapping, neural circuits [46]. However, if the learnt behaviour improves fitness, then this would increase the selection pressure on any innate elements of that learning as well as the other physical traits. This observation sets an important criterion when considering a simulation of these processes: the phenotype and its resultant fitness need to be a product of both the evolution and learning with the learning being able to guide the evolution and the evolution placing constraints on the learning. Unlike the Hinton & Nolan model, and other models of plasticity discussed, previously learning and evolution should not be competing for complete control of the phenotype and should not have a simple one to one relationship.

Chapter 3

Conditions for Learning to Become Innate

3.1 Introduction

In this chapter, the necessary conditions under which in-life learning can be genetically assimilated to become innate behaviours are examined. Through the analysis of extensions to classic models of learning and evolution, we explore some of the contradictory results in the literature regarding what is needed for learning to genetically assimilate. This investigation centres on three key questions:

Firstly, is a cost of learning a necessary driver for the genetic assimilation of learning to occur? As briefly discussed in Chapter 2, it is often suggested that a cost of learning is required to drive the genetic assimilation of learnt behaviours because the cost of learning provides a selection pressure for learnt behaviours to be replaced through selection by - what are assumed to be - less costly innate behaviours. Due to the complexity of experimentally measuring a cost of learning, evolutionary theory has lent heavily on computation simulations to explore this issue. These models offer contradictory conclusions: for example (as will be discussed), whilst Hinton & Nowlan's model [49] requires a cost of learning for genetic assimilation, Borenstein et al.'s model [12] achieves genetic assimilation to the global optimum without any cost of learning. Understanding why this is the case, provides important context to the work in this thesis, as the new models presented in subsequent chapters show genetic assimilation without a cost of learning.

Secondly, does there need to be *'neighborhood correlation'* [74] for learning to be assimilated? The much-cited, but little discussed, argument put forward by Mayley [74] hypothesizes that if movement in genotype space is not well correlated with movement in phenotype space then the genetics are unable to catch up with the learning.

This question is also important to the work in this thesis, as the models presented have learning and genetic change operating in separate parametric spaces but still demonstrate genetic assimilation through a variety of causal routes.

Thirdly, does learning need to be consistent across generations for it to be assimilated? West-Eberhard [129] suggests that for learning to be genetically assimilated - or have any lasting effect on evolutionary outcomes - it needs to be consistent across generations, thereby providing a consistent target to which the genetics can catch up: if the target is moving faster than the genetic change then the genetic configuration will endlessly "chase" the moving target. As Downing [29] puts it, for learning to dominate, the environment needs to be *nearly static*. The work in this thesis challenges this assumption and creates a hypothesis that learning can help discover structural regularities in the fitness landscape and this learning can be an effective fitness signal from learning to evolution despite different learning outcomes. This idea is developed in subsequent chapters.

3.2 The Cost of Learning

3.2.1 The Fitness Cost of Learning

The fitness of a phenotype is a trade-off between the benefits of a particular trait configuration in an environment versus the cost of maintaining that configuration [77]. The cost of maintaining the configuration can include a cost of plasticity. Dewitt et al. [26] (p. 78) defined the cost of plasticity as when *"in a focal environment a plastic organism exhibits lower fitness while producing the same mean trait value as a fixed organism"*. The cost of plasticity has the potential to completely offset any optimizing effect of plasticity (i.e. the higher fitness configurations facilitated by adaptive plasticity are cancelled out by the cost of expressing or maintaining that plasticity).

In the majority of models of learning and evolution, a cost of learning is the key driver of selection for innate behaviours to replace the learnt behaviours. Whilst there has been significant difficulty in experimentally detecting the cost of developmental plasticity [5] and where detected it has been difficult to quantify[1], Snell-Rood [109] suggests that the cost of learning is more clearly identifiable than other forms of plasticity. West-Eberhard [130], Mayley [74], Turney [117] and others suggest that trial-and-error learning may be particularly costly. For example, predator avoidance is a good example of a high learning cost, where time lost learning to avoid predators - rather than an immediate instinctual behaviour - may lead to a swift loss of future fecundity. Ackley &

[1]Whilst much work has been dedicated to experimentally quantifying the costs of plasticity, in Murren et al's review of evolutionary constraints to plasticity [84], they were unable to find reliable experimental data supporting an accurate quantification of costs of plasticity

Littman's [1] observations from their Artificial Life simulation supports this intuition: they observed that predator avoidance behaviour primarily evolved to become innate whereas food selection evolves to be a learnt behaviour.

It should be remembered that the plasticity associated with innate behaviours also has a fitness cost. Therefore, before discussing the detail of learning costs in evolutionary models, it is useful to compare the costs of learning to the costs of innate behavioural plasticity. Here, Auld's [5] framework for plasticity costs is used to compare the likely differences between the fitness costs of innate behaviour and the fitness costs of learnt behaviour, and this is summarized in Table 3.1 (column 3). From the table, it can be seen that both innate behaviours and learnt behaviours are likely to have a fitness cost (as well as a fitness benefit) and these costs are generally likely to be higher for learning than they are for innate behaviours. However, these conclusions are built mostly on intuition and are not strongly supported by empirical evidence. Despite this, the majority of models of learning and evolution are built on the premise that learning has a fitness cost and innate behaviours are effectively cost-free and that this cost differential is a necessary component to provide a selection pressure to drive the evolution of the genetic prescribed innate behaviours as a replacement of the learnt outcomes.

1. Type of Plasticity Cost	2. Basis of Cost – General Plasticity	3. Comparison of Fitness Cost: Learnt Behaviour vs Innate Behaviour
Maintenance	Maintaining environmental sensing or regulatory mechanisms	Bigger brains needed for learning are metabolically expensive to maintain therefore learning is likely to have an inherent maintenance cost higher than for innate behaviours [84, 110].
Production	Expressing the trait plastically (delta to fixed development cost)	Learning expends energy and takes time that would be fully productive if the behaviour was innate. Therefore during learning there is a temporary mismatch between behavioural phenotype and environment. Learning may also delay reproductive fecundity [74, 84].
Information Acquisition	Sensing the environmental input	Both learning and innate behaviours require sensing of the environment. Correct sensing of environmental input is critical to effective learning and requires increased neural input [110].
Developmental Instability	Mismatch to the environment due to plastic expression	Whilst learning is generally thought to be adaptive, it may be unreliable where learning the wrong thing has a fitness cost [74, 110]. However, innate behaviours may be maladaptive when environment change is fine-grained which has its own fitness cost. [110].
Intrinsic Genetic	Negative fitness effect of linkage between loci impacting plasticity (e.g. via pleiotropy)	As innate behaviours become increasingly genetically prescribed, there is potential for pleiotropic effects, which may have negative fitness consequences. Since, learning may be constrained by innate behaviours (see Section 2.1.2), any negative fitness effects due to pleiotropy may also accrue to learning.

TABLE 3.1: Adapting and building on the work of Auld's [5] interpreting DeWitt [26]. Of the five types of general plasticity cost identified by DeWitt (Column 1), the basis of the cost as described by Auld (Column 2) enables a comparison of the fitness costs of learning as opposed to the fitness costs of innate behaviours (Column 3)

However, there are models of learning and evolution that do not normally include a cost of learning, for example, models by Borenstein et al. [12], Saito et al. [101], and Frank [36], and importantly also the models presented in this thesis. By the addition of a cost of learning to Borenstein et al's highly respected model of learning and evolution [12], it is shown that whilst a cost of learning can drive genetic assimilation, it is not a necessary condition and can inhibit reaching globally optimal fitnesses that would be attained without a cost of learning.

3.2.2 Representation of the Cost of Learning in Evolutionary Models

From a modelling of learning perspective, the potential fitness costs of learning identified in Table 3.1, as well as the discussion by Mayley [74] suggests that three broad categories of cost should be considered:

1. Costs that are fixed no matter how much learning takes place (a fixed cost). An example being the cost associated with acquiring information.
2. Costs that are proportional to the amount of learning that takes place (a production cost)
3. Costs of any mismatch between environment and phenotype after learning (a mismatch cost).

Most models of learning and evolution that include a cost of learning calculate this cost as a factor of the amount of learning that takes place - i.e. a production cost. The cost of the mismatch between the phenotype and the environment is naturally captured by the model's fitness function, although this normally does not account for the longer period of lower fitness whilst time is taken to perform the learning (Mayley [74] categorises this as *"time-wasting costs"*). Of the models reviewed in Chapter 2 that include an explicit cost of learning, all include a production cost, none include a fixed cost for learning and mismatch cost is encoded in the fitness function.

One model by Morgan, Suchow & Griffiths [83] does not explicitly include a cost of learning but does include an implicit "developmental instability" cost. Based on the Hinton & Nowlan model, it considers the potential mismatch between the plastically expressed phenotype and the environment as a driver for assimilation. They do this by using the assumption that plastically expressed traits have a probability (p) of matching the environment whereas fixed traits always acquire the correct trait. Therefore selection will favour fixed traits. However, whilst the assumption that fixed (genetically specified) traits will evolve the adaptive trait in a fixed environment, this would not necessarily hold where the environment changes regularly.

3.2.3 The Action of Plasticity Costs in Evolutionary Models

3.2.3.1 Learning Changes the Fitness Landscape

Before assessing if a cost of learning is a necessary condition for genetic assimilation it is worthwhile considering the effect that learning can have on the fitness landscape and how a cost of learning changes that effect. Borenstein et al. [12] model learning as a hill-climb of the phenotypic configuration to the nearest local optimum giving an individual a locally optimal fitness. In keeping with Borenstein et al.'s results, Panel A of Figure 3.1 confirms this representation of learning creates plateaus of equal fitness, thereby flattening the fitness landscape. This can hide genetic variance from selection - what Mayley [73] termed the *'hiding effect'* of learning.

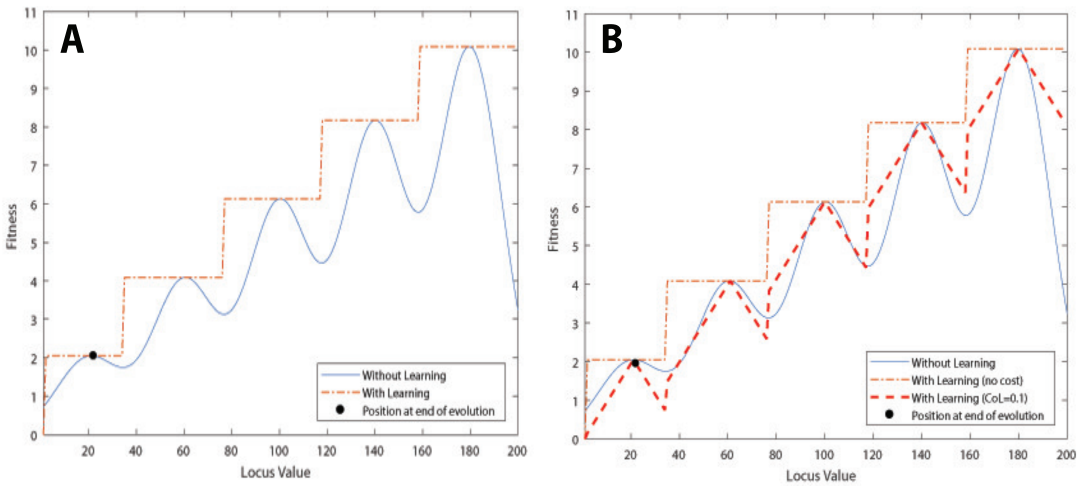


FIGURE 3.1: **A** Adapted from Borenstein et al.'s model of random work in a multi-peaked fitness landscape [12]. A single locus taking values between 0 and 200 with the fitness value being calculated as a function of stacked Gaussian functions[2]. Borenstein's hill-climb learning has the effect of flattening the fitness landscape. **B** When a cost of learning is applied based on a function of the amount of learning required, a selection gradient to the local optimum is reapplied

3.2.3.2 Methods for Calculating the Cost of Learning

As discussed in Chapter 2, the original Hinton & Nowlan (H&N) model [49] used a variable cost of learning, based on the number of learning steps, to provide a gradient for selection to reach the global optimum in the single-peaked fitness landscape:

[2]To replicate Borenstein et al's results, the following fitness landscape was used:

$$F_{Gaussian}(x) = \sum_{j=1}^{d} a_j exp\left(-\frac{(x-b_j)^2}{2c^2}\right)$$

where x is the trait value, d is the number of peaks, **a** is a vector of heights of each peak, **b** is a vector of the centrepoint of each peak and c is the Gaussian RMS width.

the closer the genetic configuration is to the fitness peak, the less learning is required, the lower the costs of learning to reach that peak, the higher the overall fitness of the phenotype, and the more likely that phenotype is to be selected. Subsequent models with a similar structure to the H&N model [8, 34] also use similar methods to provide a gradient for selection to follow.

Figure 3.1 (B) shows that a cost of learning can reapply a fitness gradient, effectively removing the hiding effect of learning.

Whilst implementing learning in different ways, other models of learning also deploy this logic to cause a selection pressure to drive the genetic assimilation of learning. For example, Ancel's [3] model of general plasticity uses a different method to simulate the phenotype's plastic response and therefore deploys a different technique to apply a cost to plasticity. In this model, the plasticity is specified as an upper and lower bound of plastic response *"an interval of phenotypic possibilities"* and evolution has the potential to expand or shrink this interval. The cost of plasticity is derived as a proportion of the length of the interval of plasticity and therefore there is a fitness benefit to reducing the range of the plastic response. Consequently, there is also a selection gradient to reduce plasticity.

3.2.4 Borenstein's Model of Learning and Evolution

Unlike most similar models, Borenstein et al.'s [12] mathematical analysis and supporting simulations do not include a cost of learning for the genotype to reach the global optimum in a multi-peaked landscape. It is this counterpoint to the general view that a cost of learning is necessary to drive genetic assimilation that makes Borenstein's model an interesting framework with which to explore the action of a cost of learning.

Borenstein et al.'s [12] model of learning and evolution uses non-symmetrical fitness proportional random walk analysis to model the population mean fitnesses in different fitness landscapes. This random walk simulates the action of selection, variation, and inheritance on the population. So, during evolution, the probability of a gene moving up or down in trait value is a function of the fitness gradient, where at trait location i, the probability p_i of taking a +1 step is $p_i = 1/(1 + e^{(F_i^- - F_i^+)/T})$ and the probability q_i of taking a -1 step is $q_i = 1/(1 + e^{(F_i^+ - F_i^-)/T})$, where $F_i^+ = F(i+1)$ and $F_i^- = F(i-1)$ and T is the Boltzmann scaling factor set at 0.1.

Learning is represented as a simple hill-climb to the nearest local optimum and whilst not specifically defined in Borenstein et al.'s paper, this has been interpreted as follows: Starting at learning step $s = 0$, where $\mathbf{g}$ is a vector representing the genome and $\mathbf{l}^s$ represents the learning phenotype l at learning step s, at the start of learning $\mathbf{l}_i^{s=0} = g_i, \forall i$, and for each learning step s:

$$\mathbf{l}_i^{s+1} = \begin{cases} \mathbf{l}_i^s + 1, & F(\mathbf{l}_i^s + 1) \geq F(\mathbf{l}_i^s) \\ \mathbf{l}_i^s - 1, & F(\mathbf{l}_i^s + 1) < F(\mathbf{l}_i^s) \end{cases}$$

where, as before, $F(\mathbf{l})$ is the fitness of the phenotype $\mathbf{l}$ after learning. Learning completes when moving up or down does not increase fitness: $(F(\mathbf{l}_i^s) \geq F(\mathbf{l}_i^s + 1) \wedge F(\mathbf{l}_i^s) \geq F(\mathbf{l}_i^s - 1)) \wedge \neg(F(\mathbf{l}_i^s) \leq F(\mathbf{l}_i^s - 1))$. Where more than one dimension is modelled, at each learning step a locus is chosen at random with learning continuing until there is no increase in fitness.

Borenstein et al.'s analysis primarily focuses on *'draw-down'*, where draw-down is the maximum loss of fitness that must be incurred to traverse the fitness valley in the multi-peaked landscape (see Borenstein et al [12] for detail). Borenstein's results clearly show that, without learning, where there is a significant gradient to the local optima, the draw-down depth is a driving factor for the number of generations to reach the global optimum. This is because there has to be a sequence of low-probability steps to escape a local optimum and assuming a consistent number of steps to cross the valley the steeper the gradient, the lower the probability of those steps. Where learning flattens the effective landscape, the draw-down is effectively removed (as per Panel A of Figure 3.1) and so the movement in genotype space is essentially genetic drift. Consequently, as will be shown below, in the learning case, the number of steps across the valley dictates the number of steps to reach the global optima: the wider a valley the longer it takes to drift to the next local optima.

3.2.5 Genetic Drift is Effective at Catching-up With Learning

In a reproduction of Borenstein et al.'s original model and using a two-trait Rastrigin landscape[3], Panel A of Figure 3.2 shows that for the non-learning case, the instances of the population mean phenotype get trapped at low fitness local optimum. Panel B of Figure 3.2 shows that, when guided by learning, the population mean phenotype reaches the basin of the global optimum fitness.

To understand how the phenotype can reach the global optimum without a cost of learning providing a selection pressure, it is worth examining in detail how learning and random-walk evolution interact to reach the global peak in a multi-peaked fitness

[3]Borenstein uses a *"modified"* version of the Rastrigin landscape. However, the Rastrigin formula provided by Borenstein et al in the Supplementary Materials does not appear to yield the landscapes shown in the main text. Here the modified Rastrigin equation for the fitness landscape used consistent with the landscape shown in Borenstein's figures is

$$F_{Rastrigin}(\mathbf{x}) = - \quad C_r \cdot d + \sum_{i=1}^{d} x_i^2 + C_r \cdot cos(2\pi x_i))$$

Where C_r controls the depth of the peaks, d is the number of traits and x_i is the trait of either the genotype or phenotype and can take the values $x_i \in \{-5.0, -4.8, ..., 0\}$

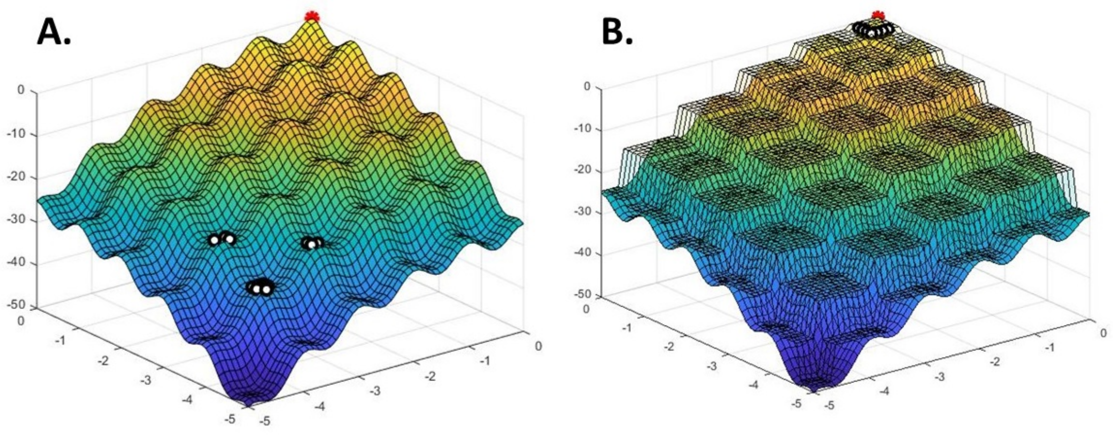


FIGURE 3.2: Replicating Borenstein et al.'s results from a model of random work in a multipeaked fitness landscape [12]. A dual locus example taking values between 0 and -5 with the fitness value being calculated according to a Rastrigin function. **A.** Without learning at the end of 1,000,000 evolutionary episodes the phenotypes (black-edged circles) are stuck in local optima. **B.** When hill-climb learning is applied, the phenotype (black-edged circles) can reach the same plateau as the optimum fitness (red star) within an average of 800 evolutionary episodes (30 runs).

landscape. As stated by Borenstein and shown in Figure 3.1, the *'ideal learning'* (learning which moves the phenotype to the nearest local optimum and therefore the population mean individual gains locally optimal fitness) flattens the *'effective landscape'* into a series of steps. As evolution is modelled as a random walk, whilst on a plateau, the value of a gene has an equal probability of going up or down; the genotype can essentially drift across the flattened basin of attraction for each local optima. When this drift encounters the boundary to the next basin of attraction (the bottom of the step shown in Figure 3.2), then through learning the phenotype can jump to the next basin of attraction and the genotype will then drift across the next flattened peak. The genotype is extremely unlikely to fall off this plateau as this would entail a large drop in phenotypic fitness (which has a very low probability of occurring in the probabilistic random walk). This provides a ratchet to higher fitnesses, so the movement in genotype space is not a pure form drift - it is a *ratcheted drift*. This mechanism is illustrated in a reproduction of Borenstein et al.'s results[4], provided in Panel B of Figure 3.2, the movement of the genotype from one peak to the next relies on genetic drift followed by jumps in fitness. As demonstrated by Borenstein et al.'s results, this drift has a higher probability of reaching the next peak than on an unflattened valley (i.e. without learning).

[4]The exact formula for producing the two-dimensional fitness landscape is not in Borenstein et al.'s paper and consequently there are small differences in the actual result which are not material to the point being made.

3.2.6 Drift Effectively Assimilates Learning Over Multiple Traits

One may expect that drift may not be an efficient mechanism for the genetics to catch-up with the learning, especially where the learning impacts multiple traits and therefore needs to drift in multiple dimensions. However, as shown in the simulation results using an extension to Borenstein et al.'s model in the Rastrigin landscape (Figure 3.3), the time for the phenotype to reach the global optimum (which is the same as the time for the genotype to drift to the top plateau in the fitness landscape) - enabled by learning - scales quadratically with the number of traits (dimensions of the fitness landscape). Therefore the genotype has the potential to catch up where learning impacts multiple traits.

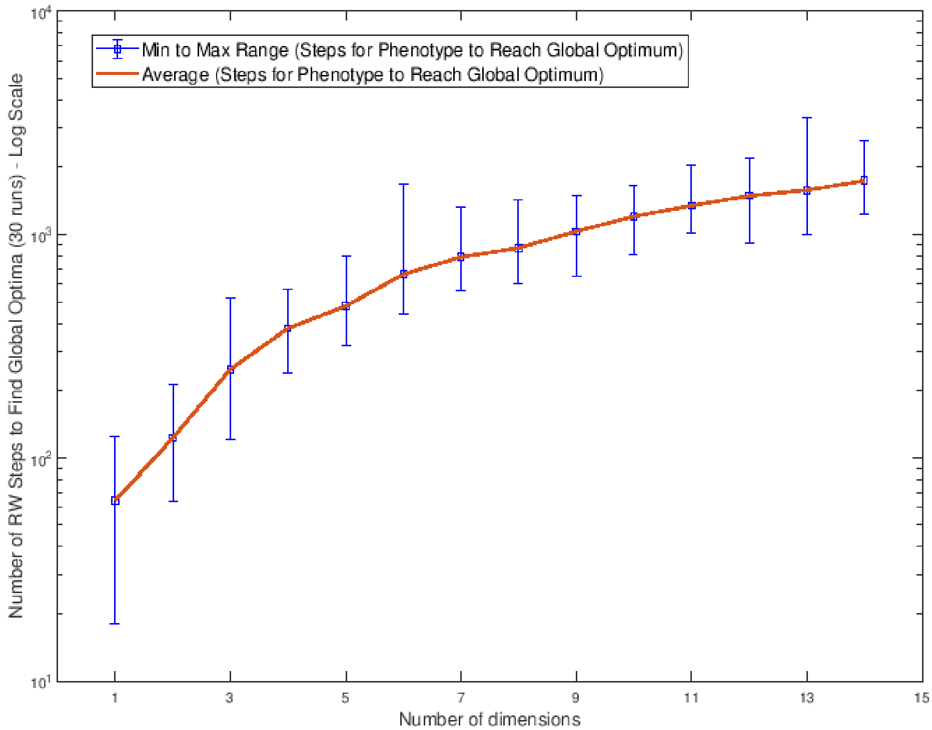


FIGURE 3.3: Analysis of the average time for the phenotype to reach the global optimum with learning. For 30 runs, the minimum, maximum and average times for the phenotype to find the global optimum fitness with learning and probabilistic random-walk evolution are shown plotted for a varying number of traits. For this assay, a Rastrigin landscape is used with a varying number of dimensions where each trait is a dimension $x_i \in \{-5.0, -4.8, -4.6, ..., 1\}$, $d \in \{1, 2, ..., 14\}$, $C_r = 2$

It should be noted that the time for the genotype to encounter the global optimum configuration, completely catching up with learning (complete assimilation), scales exponentially with the number of trait dimensions (not plotted). However, since there is no selection pressure for a genotype with the optimal configuration to stay at that optimum; it will likely drift away, and therefore the time for the genotype to encounter the optimum has limited relevance.

Importantly, genetic drift is not a feature specific to Borenstein et al.'s model of evolution and learning. Drift is a feature of most evolutionary models, where a mutation is retained if it provides the same (or greater) phenotypic fitness. Therefore, one would expect drift can enable genotypes to reach a global optimum in a multi-peaked landscape under learning to be generally applicable. However, whilst hard to quantify, most evolutionary theorists would contend that there is a cost to learning and from Table 3.1 it would seem this cost is likely to be greater than the costs associated with innate behaviours. Consequently, drifting innate behaviours catching up with learnt behaviours may not be a biologically likely scenario.

3.2.7 A Cost of Learning Will Inhibit Reaching the Global Optimum with Local Learning

As discussed earlier and as shown in Figure 3.1, a production cost of learning effectively reapplies the selection gradient towards the local optimum where learning is a hill-climb to a local optimum. This would suggest that, where there is a probabilistic random-walk simulation, the cost of learning would inhibit the genotype from reaching the global optimum because the phenotype is constrained to a local optimum.

To check this is indeed the case, Borenstein et al.'s model of learning and evolution has been extended to include a cost of learning. For convenience, in this implementation, the cost of learning c is calculated as the Euclidian distance[5] between the learned phenotype and the genotype (assuming the phenotype without learning is the same as the genotype); this representing the amount of additional 'work' that learning needs to do to reach the local optimum and like other models of learning and evolution, is a performance cost. This distance is scaled using the factor ϕ, so the overall cost of learning is:

$$c = \phi \sqrt{\sum_{i=1}^{n} (p_i - x_i)^2} \tag{3.1}$$

Where n is the number of loci in the genome, p_i is the phenotype trait value at loci i and x_i is the genotypic trait value at loci i.

Figure 3.4 shows that for the majority of scale factors (ϕ), a cost of learning inhibits reaching the global optimum and for values of ϕ above 0.05, the number of random-walk steps to reach the global optimum scales exponentially with ϕ. This result reflects the intuition that a cost of learning, by increasing the selection gradient towards the nearest local optimum, would increase the time to find the global optimum as compared to no cost of learning as it increases the attraction of the local optima. However,

[5]This is similar to the cost calculation in Draghi & Whitlock [31] although they use more sophisticated scaling

as can be seen in Figure 3.4, there is a small range of values for ϕ where the cost of learning does not seem to influence the speed at which the local optima is found. With a cost of learning scaler of 0.05, the overall cost of learning is a fractional percentage of the range of fitness values that the phenotype can take: the average cost of learning is 1.3% of the average phenotypic fitness value. Whilst it is difficult to draw any biological conclusions concerning the cost of learning (especially as a cost of learning is so difficult to measure), what is demonstrated is that for a large range of values, the cost of learning inhibits reaching the global optimum.

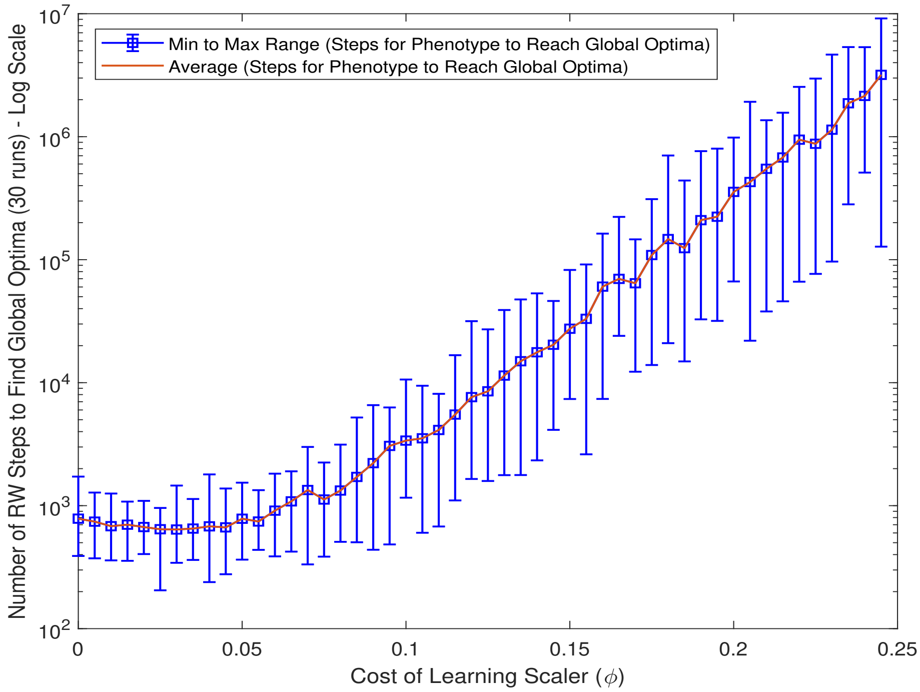


FIGURE 3.4: Using an extension of Borenstein et al.'s model of learning and evolution in a multi-peaked landscape, and the number of random-walk steps needed to find the global optima assessed $x_i \in \{-5.0, -4.9, -4.8, ..., 1\}$, $C_r = 2.$ for a variety of cost of learning scalers, where the cost of learning is based on the Euclidian distance between the phenotype and genotype

Again, whilst the action of a cost of learning has been investigated using Borenstein et al.'s probabilistic random-walk model of evolution, one would expect the same results in other models of evolution where learning moves the phenotype towards local optima. Where learning discovers the nearest local peak in a fitness landscape, a cost of learning reintroduces a gradient to the local peak on which a population will normally get stuck.

3.2.8 A Cost Learning can Enable Genotypes to Cross Fitness Valleys

In the previous sections, it was demonstrated that genetic drift induced by the 'hiding effects' of learning can allow the genotype to reach the globally optimal fitness but that a cost of learning can inhibit this process by reintroducing a selection gradient to a local optimum. However, in these scenarios, learning is constrained to finding the nearest local optimum. There are models of learning and evolution where learning can traverse fitness valleys, moving the learned phenotype to a different local optimum which potentially can eventually be the global optimum. For example, in Section 3.3.2 of this chapter, using a reimplementation of Mayley's [74] model of learning and evolution, learning can change several trait dimensions in the same learning step and, through a process of hillclimbing, can move towards the global optimum. Under these circumstances, a cost of learning can have an altogether different effect, speeding up convergence to the global optimum.

To briefly demonstrate this, consider the extreme case where learning is able to reach the global optimum, no matter what the genetic configuration is. As shown in Figure 3.5, with this type of extreme learning, the fitness landscape with learning becomes a single plateau but a cost of learning reintroduces a gradient.

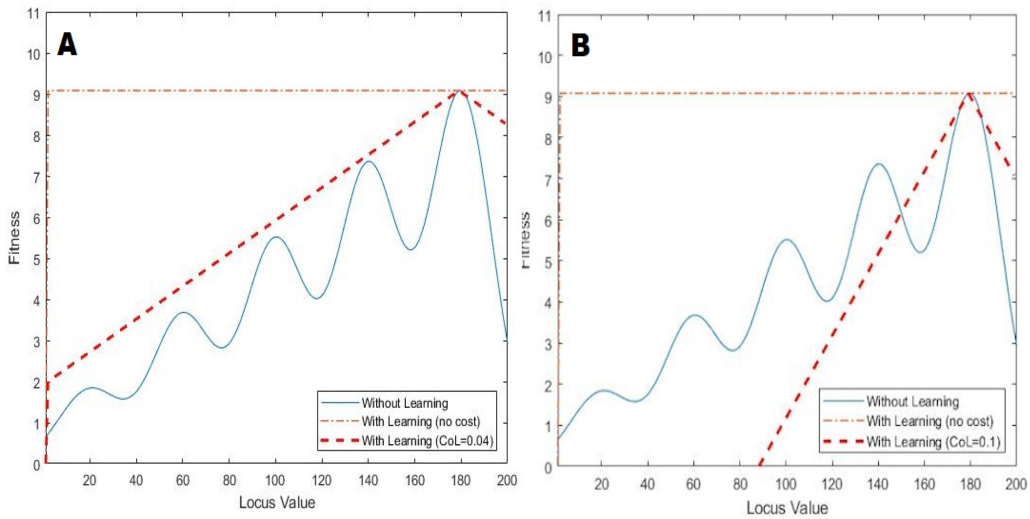


FIGURE 3.5: Adapted from Borenstein et al.'s model of random work in a single-peaked fitness landscape with learning that always finds the globally optimum phenotype. With a cost of learning (orange line) the fitness landscape is a single plateau of optimum fitness. Without a cost of learning (red dotted line), there is a consistent gradient to the global optimum (model as per previous figures). Panel A with a low cost of learning $\phi = 0.04$ and Panel B with a high cost of learning $\phi = 0.1$

The results shown in Figure 3.6 confirm that the higher the cost of learning, the steeper the gradient to the fitness optima and the faster the population mean phenotype will reach the global optimum. All costs of learning are better than relying on genetic drift (i.e. where the cost of learning coefficient ϕ is zero).

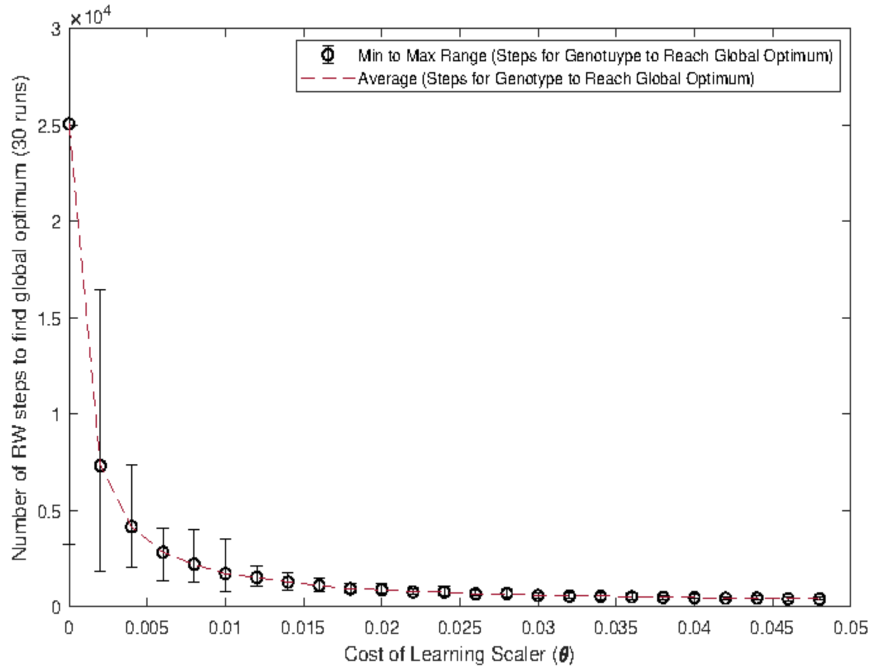


FIGURE 3.6: As per Figure 3.5 where learning always finds the globally optimum phenotype. The number of random walk steps to reach the global optimum is plotted for various costs of learning.

It is also worth noting this extreme case of learning, it is possible that for higher costs of learning coefficients, the cost of learning outweighs the benefits (see Panel B of Figure 3.5). This constrains the range of values that are sensible for this coefficient.

3.3 Neighbourhood Correlation

3.3.1 The Concept of Neighborhood Correlation

In an often-cited paper, Mayley [74] proposes the concept of *'neighbourhood correlation'* which potentially has great significance when considering the genetic assimilation of learning. Neighbourhood correlation describes the extent to which the distance moved in genotype space is correlated or uncorrelated with the distance moved in phenotype space. Mayley claims that, even where evolution and learning are working on the same task, there needs to be a high degree of correlation for genetic assimilation to take place:

> *"It is also noted that genotypic space, in which evolution operates, and phenotypic space, in which adaptive processes (such as learning) operate, are, in general, of a different nature. To guarantee that an acquired characteristic can become genetically specified, these spaces must have the property of neighborhood correlation,*

> *which means that a small distance between two individuals in phenotypic space implies that there is a small distance between the same two individuals in genotypic space."*

Mayley further defines neighbourhood correlation as the closeness of two measures: a genetic distance calculated by the Hamming distance between two genotypes and a phenotypic distance defined by the number of loci flipped within a learning operation (the L-neighborhood)[6].

Mayley used Figure 3.7 to illustrate how changes in genotype space can be correlated or uncorrelated to movements in phenotype space by examining the 'distance' the genotype will need to move to assimilate the phenotypic fitness difference delivered by learning. In the left-hand side of the figure (a), p2′ will be selected in preference to p1′ because its fitness is higher and, because the movement from g2 to g2′ is similar to the distance in phenotype space (p2 and p2′), there can be genetic assimilation. However, on the right-hand side of the same figure, genetic assimilation is less likely to occur because even though p2′ will be selected in preference to p1′, there is a large difference between the movement in genotype space (g2 and g2′) and phenotype space (p2 and p2′).

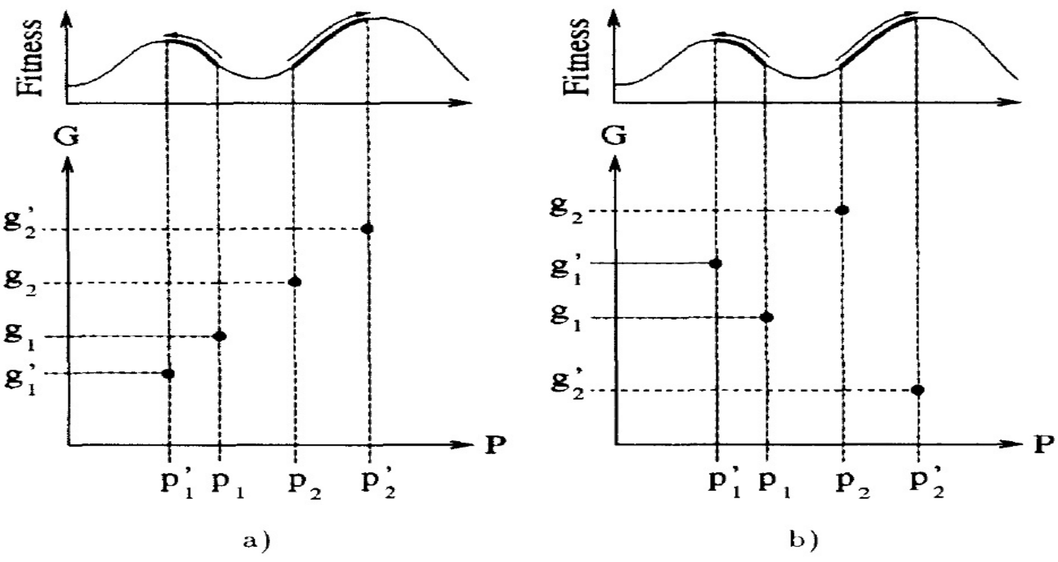


FIGURE 3.7: Reproduced with permission from Mayley [74]: a) The movement in phenotype space towards a higher fitness (p1 to p1′ and p2 and p2′) is correlated with the distance in genotype space (g1 to g1′ and g2 to g2′). b) The movement in phenotype space towards a higher fitness (p2 and p2′) is uncorrelated with the distance in genotype space (g2 to g2′).

[6]Mayley's analysis suggests that there is enough of a correlation between the number of learning operations and the movement in genotype space to consider the parameter L to be a good approximation for the degree of neighborhood correlation (see Section 5.4 of Mayley's paper [74]).

3.3.2 Mayley's Model for Testing Neighborhood Correlation

Mayley tested his hypothesis using a computational simulation of a small population of individuals that could learn and were subject to evolution. The fitness of the *'learned phenotype'* was assessed according to an NK landscape as developed by Kauffman [56]. The NK landscape method allowed Mayley to control the ruggedness of the fitness landscape with parameters: for a given number of traits (N), the larger the number of epistatic interactions (K), the more local maxima generated in the landscape.

Mayley implemented learning as a multiple bit-flip hill-climber, where one parameter (L) controlled the number of phenotypic loci flipped during a learning trail. Starting by setting the phenotype to be equal to the genotype's configuration, one learning operation involved performing N trials of L loci bit-flips to the phenotype, where the fittest trial after learning was retained and became the start-point for the next learning operation. The hill-climb proceeded until none of the L-neighborhood (flipping L loci) provided an increase in fitness. Mayley called the defined test changes to the phenotype that were used to generate each learning trial, the L-neighbourhood. Interesting, Mayley chose to implement this mechanism by defining the L-neighborhood *"at the beginning of a program run"* and therefore all generations have the same intrinsic learning abilities (which traits can change in a learning operation), although the outcome of this learning will depend on the initial genetic configuration. However, as will be shown in the next section, the pre-definition of this learning ability has a qualitative impact on the results.

To model evolution, Mayley used a small sexual population of fifty individuals and a genome of twenty genes with point mutations, rank-based selection, and cross-over of each locus based on a parameterised probability. The probability of each loci being mutated was set as $\frac{1}{N}$, so on average one loci would be mutated each generation. A cost of learning was calculated as a factor (c) of the number of learning operations and was used to create a selection pressure for genetic assimilation. So that cross-over did not create large movements in genotype space, at the start of evolution, the population was 'converged' on a single random genotype with a 5% chance of any individual's trait being different from the rest of the population.

Mayley claims his results show that large movements in phenotype space can prevent genetic assimilation. However, from the data presented by the author, it is unclear what is driving the results. The random nature of the fitness landscape and the lack of comparison to the globally optimal fitness value make it difficult to ascertain if simulations are getting stuck in local optima, or are crossing fitness valleys or reaching a local optimum at all. In addition, whilst learning's degree of movement in phenotype space is controlled by L[7], the number of mutations per generation is not well controlled

[7] In Mayley's implementation rows in the L-neighbourhood table are distinct and so multiple flips of the same loci are not available during a single learning operation.

as it can vary between zero and the number of loci, dependent on chance. The results data presented shows learning offers a small increase in population fitness at the end of evolution for rugged landscapes (where $K \neq 0$), although this effect is very small (between -1% and 7% when $K = 5, 10, 15$, c = 0.015 and $L = 1, 2, 5, 10$). With cost-free learning (c=0), the results also indicate that as the ruggedness (K) of the landscape increases, there is an increased fitness due to learning, but the degree of increased fitness reduces as L (the lack of correlation between genotype and phenotype space movement) increases.

However, whilst the results are not clear in the rugged landscapes, the results in the smooth fitness landscape ($K = 0$) with five traits flipped per learning step ($L = 5$) best show how the mismatch in movement in genotype space and phenotype space may prevent assimilation even where there is a cost of learning. As shown in a reproduction of Mayley's original result in panel C of Figure 3.8, over 20 runs, there is still on average one learning step taken by each individual in the population. It is this apparent lack of complete assimilation that is the focus of the additional analysis of Mayley's work presented in the next section.

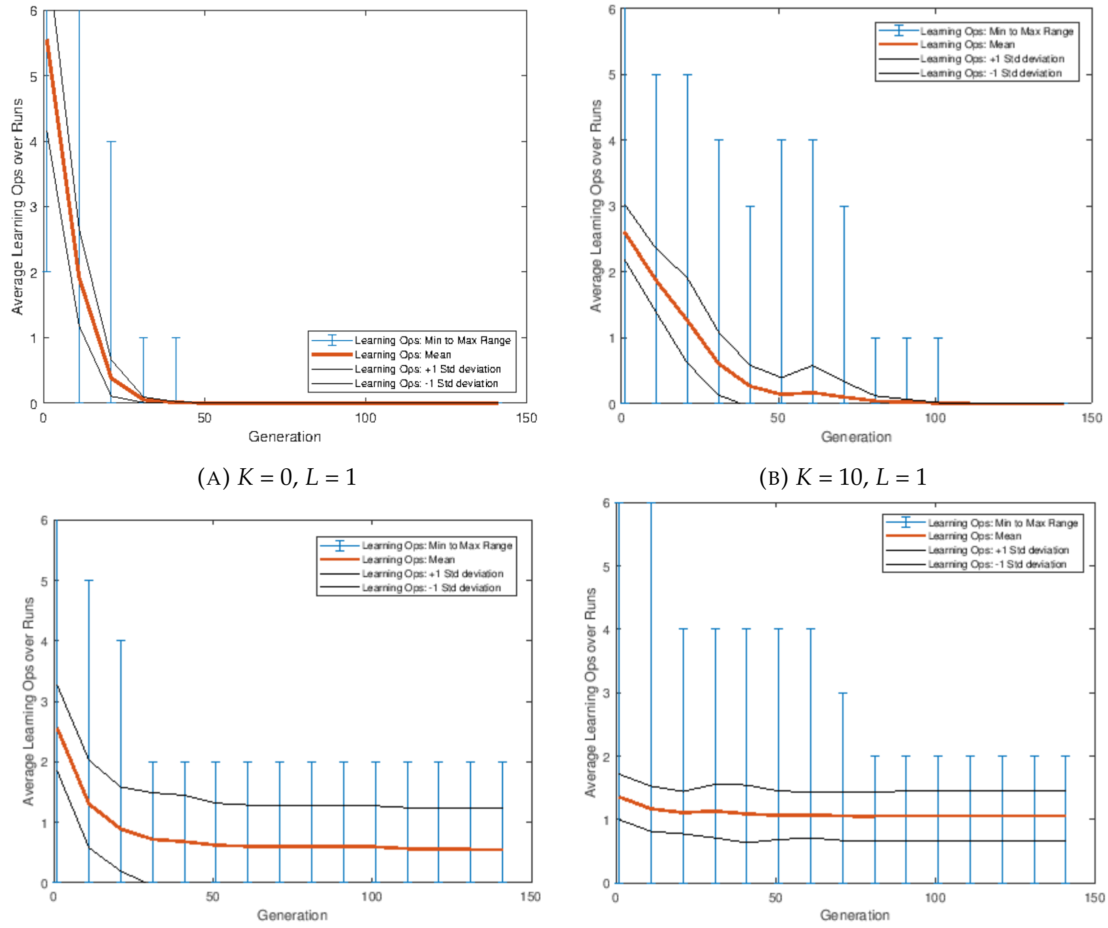
3.3.3 Examining Mayley's Result

Please note that the additional analysis of Mayley's results presented in this section is in no way meant to be a criticism of Mayley's important work. The original paper was written in 1997 when computing resources were significantly poorer than they are today and some of the modelling options explored here may not have been available to Mayley at the time.

First, we verify the operation of the extended version of Mayley's model by showing reproducing the results in Mayley's paper. This ensures that results from the extensions to the model are based on a common set of implementation details.

As per Mayley's hypothesis, as shown in Figure 3.8 (A), where the movement in phenotype space (L=1) is similar to the movement in genotype space (on average one mutation per generation) then genetic assimilation reliably occurs. And where there is a relatively large movement in phenotype space (L=5) as compared to movement in genotype space, there is an absence of reliable genetic assimilation; although it does on occasion occur in both the smooth and rugged fitness landscapes.

However, unlike the original result, if the experiment in the smooth fitness landscape (K=0) and a large phenotypic movement (L=5) is run for long enough, then reliable assimilation does occur, as shown in Figure 3.9 (I). This is because, whilst on average there will be one mutation per child per generation, the mutation of each allele is probabilistic and therefore there is scope, with a low probability, for the movement in genotype space to be large enough to replace the movement in phenotype space (i.e. mutation



(A) $K = 0, L = 1$

(B) $K = 10, L = 1$

(C) $K = 0, L = 5$

(D) $K = 10, L = 5$

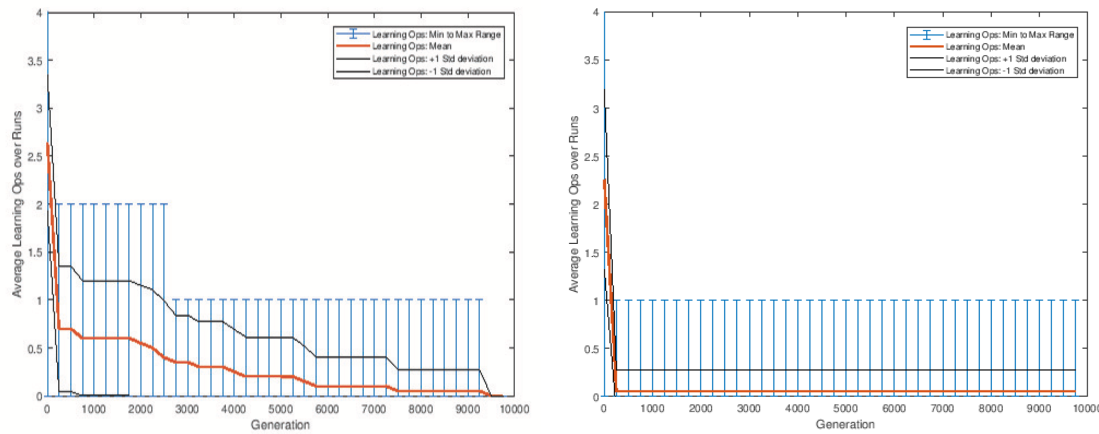
FIGURE 3.8: A reproduction of Mayley's [74] results for two different ruggedness of fitness landscapes and two different L values (number of bits flipped during a learning operation) with a low-cost of learning ($c = 0.015$). **A**) Assimilation occurs in a single sloped fitness landscape ($K = 0$) and a single bit-flip learning operation ($L = 1$), **B**) Assimilation also occurs for a rugged fitness landscape ($K = 10$) and a single learning operation. **C**) Even though the fitness landscape is a single slope ($K = 0$) assimilation does not appear to occur where there are multiple bit-flips during a single learning operation. **D**) Assimilation also does not appear to occur in the rugged landscape ($K = 10$) where there are five bit-flips per learning operation. Note the scale is that used in Mayley's original paper to aid comparison.

to change multiple alleles at once). To test this we alter the experimental setup so that the number of mutations per offspring is strictly controlled via the parameter m, which defines the number of loci mutated (flipped) at each generate where a vector of m loci to flip after inheritance is defined by $\mathbf{l}$ and where g_i is the i^{th} gene being mutated, so that:

$$\mathbf{l} = (l_1, l_2, \ldots, l_m), \mathbf{l}_i \sim U(\{1, 2, \ldots, n\})$$

and

$$\mathbf{g}_i = \begin{cases} -\mathbf{g}_i, & i \in \mathbf{l} \\ \mathbf{g}_i, & \text{otherwise} \end{cases}$$



(I) With the original experimental parameters ($K = 0, L = 5$), when the number of generations is extended to 10,000, genetic assimilation of the learning eventually occurs.

(II) For the extended number of generations, if the number of mutations per generation is strictly controlled then as per the original result, genetic assimilation of the learning does not occur in all runs.

FIGURE 3.9: A reproduction of Mayley's [74] results as per **C)** of Figure 3.8 this time extending the run to 10,000 generations with and without strict control of the number of mutations.

Figure 3.9 (II) uses the same parameters as that used to produce the results in Panel I of the same figure but with the number of mutations per individual set to one ($m = 1$), confirms that in Mayley's experimental set-up, even in a smooth fitness landscape, when strictly controlled, small movement movements in genotype space cannot assimilate where there are large movements in phenotype space ($L = 5$).

Why this should be the case, for a strictly controlled mutation rate in a smooth fitness landscape, can be understood using the following rationale:

- The parameter L defines the minimum number of loci flipped in a single learning operation and so there is also a minimum fitness effect of a learning operation.
- One learning operation may reverse one or more of the bit-flips of the previous learning operation and therefore the total number of loci flipped after learning may be less than the number of learning operations multiplied by the number of loci flipped per learning operation.
- A series of learning operations will bring the phenotype to within x 'incorrect' loci of the fitness peak.
- The fittest learnt phenotype will come to dominate the small population.

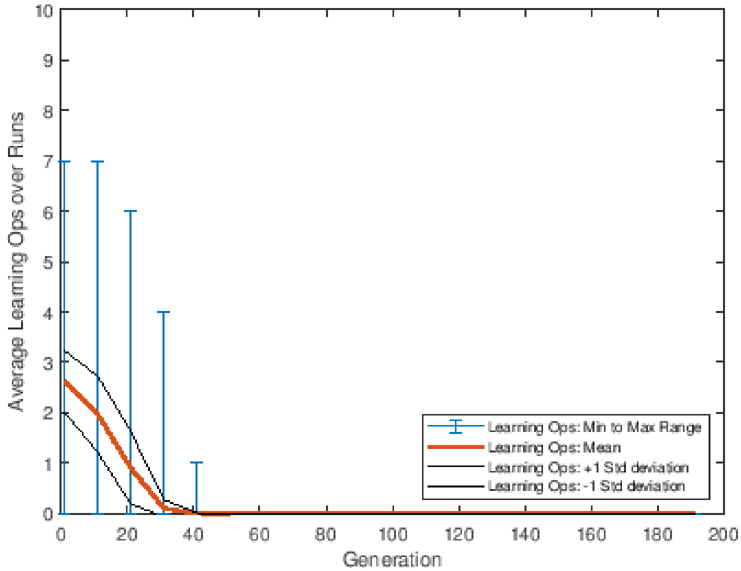
- A mutation will only fix in the population if it moves the phenotype closer to the fitness peak for the same number or fewer learning operations (i.e. the mutation reduces x or reduces the number of learning operations).
- By a certain generation, mutation will move the genotype to a start point for learning where no alternative combination of the learning explored in the L neighborhood will get the phenotype closer to the fitness optimum for the same number or fewer learning operations (this is increasingly probable the genotype gets closer to the fitness optimum and the number of learning operations decrease).
- Once at the generation where learning can move no closer to the optimum is reached - usually when the genotype is only one learning step away from the learnt phenotype - it is the high fitness differential between the large movement in phenotype space and the small movement in genotype space that inhibits the complete assimilation.

This logic is contingent on the L-neighborhood being consistent across all generations, i.e. learning available is the same from one generation to the next. This is the case with Mayley's experimental set-up. However, if the L-neighborhood changes between generations, then learning will be assimilated even when $m = 1$ because eventually there will be a combination of learning operations that enable the genotype to get within one mutation of the learnt phenotype and when that happens a subsequent mutation can replace the last learning steps. An assay where, for the same experimental parameters as the results shown in Figure 3.10, the L-neighborhood is reset at each generation confirms this to be the case.

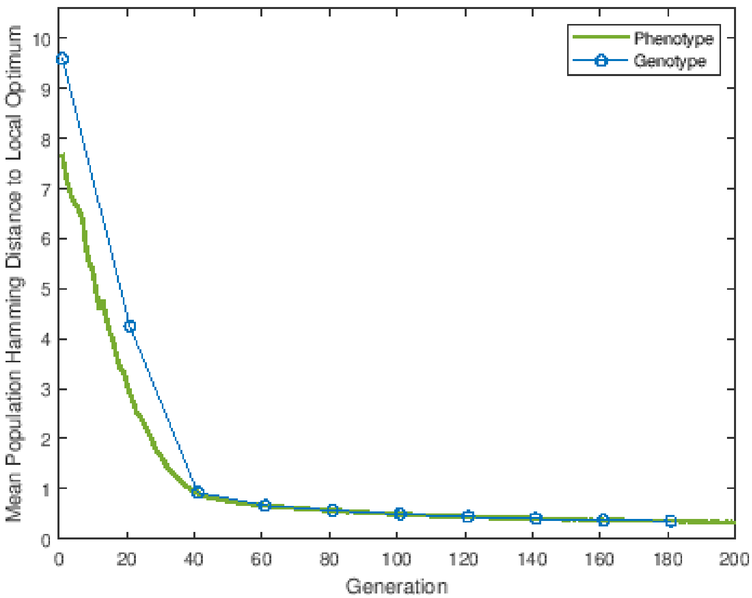
It should be noted that due to the forced mutation at each generation, once full assimilation between the genotype and phenotype has occurred, the genotype (and phenotype) will not always stay at the optimum as mutation may move it away from this optimum. This is verified in Figure 3.10 II, depicting the population average Hamming distance from genotypes and phenotypes to local optimum. This indicates that not all individuals in the population are at the local optimum.

3.3.4 The Implications of Neighbourhood Correlation

The results presented in the previous section confirm that for a single-peaked landscape, Mayley is broadly correct in saying that genetic assimilation of learning is dependent on the size of movement in phenotype space being similar to the size of movement in genotype space. A caveat to this is that neighborhood correlation is important where each learning step has a minimum step size and therefore a minimum fitness effect. This is why even though other models of learning and evolution use hill-climb learning that produces large movements in phenotype space - for example, Hinton &



(I) As per Figures 3.8 and 3.9, the average number of learning operations is plotted every ten generations.



(II) The population mean distance to the optimum fitness is tracked for the genotype and phenotype. This confirms that not all genotypes or phenotypes in the population reach optimum fitness.

FIGURE 3.10: Adaption of Mayley's model so that the L-neighbourhood (which defines combinations of traits that can be changed together) is reset at each generation, rather than being fixed at the start of evolution. In this scenario, across evolutionary time the L-neighborhood is searched more comprehensively and therefore combinations of learning operations that move the genotype to the global optimum are quickly found. All other parameters are as per previous experiments.

Nowlan and Borenstein et al. - the minimum step-size of learning in these models is one locus change, and this is equivalent to L =1 in the Mayley model: learning can always take a small step if needed so full assimilation can take place. Importantly, unlike the Hinton & Nowlan model, a model of learning that flips multiple bits allows learning to cross fitness valleys[8] creating potential for the Baldwin Optimizing Effect.

Mayley's hypothesis is dependent on there being a limit on how learning explores the phenotype space across the generations (i.e. a limited L-neighborhood). The perhaps more realistic assumption that the available learning changes over evolution time shows that learning can usually find a path to full assimilation if the search of phenotype space is unconstrained.

Mayley's neighbourhood correlation concept is of interest - especially when considering learning discovering a globally optimal fitness in a multi-peaked fitness landscape. However, when only considering neighbourhood correlation in terms of the minimum size of the learning step - as modelled by Mayley - it is not a concept that is universally applicable as perhaps it would initially appear: large total movements in phenotype space can be assimilated as long as that large movement is achieved in steps that cause a fitness difference the same or smaller than that of a single mutation (or multiple mutations where that is the evolutionary assumption).

3.3.5 Is Neighbourhood Correlation Only About Distance?

Despite the intricacies discussed above, the general concept that a mutation effect that creates a small fitness difference is unlikely to be able to replace a learning step that creates a large fitness holds true. Given the model that Mayley deploys uses a one-to-one genotype to phenotype mapping, is the same true for complex genotype-phenotype maps?

First considering relative distances, a one-to-one genotype to phenotype map means that a change in genotype allele value (in this case flipping of one locus) leads to a similar change in a phenotype's trait value (flipping one trait). However, in a complex genotype-to-phenotype map such as one created by a developmental process [126], small movements in genotype space may produce either no movement in phenotype space due to canalisation [105] or potentially, large movements due to the effects of thresholding [129]. In the case where a small change in the genotype creates a large change in the phenotype, there is an improved potential for the genotype to catch up with the learnt phenotype. Where the opposite is true, there should be a decreased probability of assimilation.

[8]Borenstein's model can cross fitness valleys without multiple bit flips because fitness-reducing steps are accepted with some probability, but this becomes increasingly unlikely as the size of the 'draw-down' increases.

But is it only relative distances in genotype space and phenotype space that determines assimilation?

Regardless of the size of movement, complex genotype-phenotype maps also may lead to accessibility issues, where the learnt outcome is not accessible through mutations. In theory, if the search of phenotype space by learning is unconstrained but the innate phenotype that can be produced by evolution is constrained, then some learnt behaviours may not be accessible and therefore cannot be genetically assimilated. Mayley [74] points to this when discussing the genotype being on one peak and the phenotype being on another. Downing [29] in the discussion of his Artificial Life inspired modelling results, appears to allude to this with the words:

> *"Essentially, our results do nothing to dispute the claim that learned traits are extremely difficult to reverse-engineer into a recipe-type genome. But then, it would require quite a leap of faith to believe that any fixed, largely deterministic generative process could map all m-bit genomes into all n-bit phenotypes (where $n \gg m$)."*

The potential for an accessibility issue for genetic assimilation in complex genotype-phenotype maps is a complex topic and has not been explored further in this thesis. Further dedicated in-depth exploration of this topic is potentially an area of future work (see Section 8.2)

3.4 Consistency of Learning

A common assumption is that the outcome of learning needs to be consistent across generations for there to be a stable target for the genetics to assimilate to. For example, West-Eberhard [129] identifies that a learning outcome needs to be repeatable for there to be potential for learning to affect physical traits and innate behaviours. This is important because one of the fundamental assumptions of the Hinton & Nowlan [49] and all other models of the Baldwin Effect is that there is a stable adaptive learning outcome across individuals in a population and the target of evolution has consistency across multiple generations.

3.4.1 Assimilation of Learning and Rate of Environmental Change

Rago et al.'s [97] analysis of the evolution of plasticity in environments that change at different frequencies provides insight into how the stability of the learning target may affect assimilation. Whilst not specifically referencing learning, the costly plasticity[9] in this model was reversible (for example, temperature-dependent colouration)

[9] In the experimental set-up, the cost of plasticity was independent from the magnitude of plastic expression

and reversible plasticity can serve as a suitable conceptualization for learning. Their experimental set-up explored the evolution of plasticity in fine-grained and coarse-grained environments, where fine-grained environments are those in which the evolutionary target changes within an individual's lifetime and course-grained environments change the target of selection on a timescale that spanned multiple, non-overlapping generations. Similar to that used by Lande [63], the classic linear reaction norm method was deployed, where the intercept (breeding value) and slope (direction and magnitude of plasticity) are both subject to genetic variation.

In common with Lande's [63] result, Rago et al. found that plasticity evolved in rapidly varying *fine-grained* environments but did not in more slowly changing coarse-grained environments. Crucially, in these *coarse-grained* environments where plasticity did not evolve, there was an initial plastic response on the change of environment that was quickly removed by evolution (suggesting assimilation of the plastic response - what Rago et al. term *'plasticity minimization'*). In most coarse-grain environments, adaptive plasticity did not evolve but the authors found that in a fast coarse-grained environment - one where there was the environment changed at every generation but individuals only experienced one environment - plasticity also evolved, suggesting that multiple environments within a lifetime is not a necessary condition for costly plasticity to evolve. Relating these results to West-Eberhard's intuition: a rapidly changing environment favours the evolution of reversible plasticity (e.g. learning) but the frequency of change of the adaptive peak (found through learning) means that there is little scope for slow-changing genetics to catch-up. In a slow-changing environment, learning may not get the chance to evolve because the evolution of the genome can adequately track the change in environmental conditions.

A more specific learning based example of the effect of a change to the target learnt phenotype has been provided by Christiansen et al. [18] through simulation designed to test for the evolution of a *Universal Grammar* in human language. This model is of special interest because language evolution is often thought of as a case where the Baldwin Effect is likely to be observed. As neatly summarised by Morgan, Suchow & Griffiths [83], this is because the fitness benefits of language are thought to require more than one individual and therefore language is more likely to arise and spread through plasticity rather than mutational effects[10]. Potential evidence for this includes plasticity in birdsong development [83].

The Christiansen et al. model was based on a structure similar to that of the classic Hinton & Nowlan model [49], where genes can be zeros or ones or plastic as represented by a question mark, although 'fixed' traits may be expressed differently with a low probability. The target language was specified similarly as a set of binary features with one gene representing each feature of the language. Learning was modelled as

[10]Morgan et al also suggest there may be a more limited case for the evolution of language where language arises through genetic mutation and then spread through plasticity.

a sampling of changes to the phenotype based on biases in the genotype (in this case 95% for fixed traits and 50% for learning traits). The fastest learners, those reaching the target language specification, were selected for the next generation, where evolution was modelled as a random crossover between two individuals. In contrast to the Hinton & Nowlan model, the target features of the language are allowed to change over evolutionary time. Importantly, in this model, the genes were used to constrain how the specification of the language changed; with p percentage of the change determined by high-frequency genes in the previous generation and determined by the language $1 - p$ percent of the time. To reflect that language can change due to cultural evolution, a random change of ten times that of the genetic mutation rate was also introduced into the model. The simulation results presented by Christiansen et al. show that unless there is a significant genetic constraint on the rate that the target language evolves (e.g. $p > 50\%$), the genetics are unable to catch up with the fast-evolving language. These results were further summarised by Heyes, Chater and Dwyer [48] with a figure, not in the original Christiansen et al. paper, reproduced in Figure 3.11 that compared the extent to which plastic genes became fixed for different rates of language change.

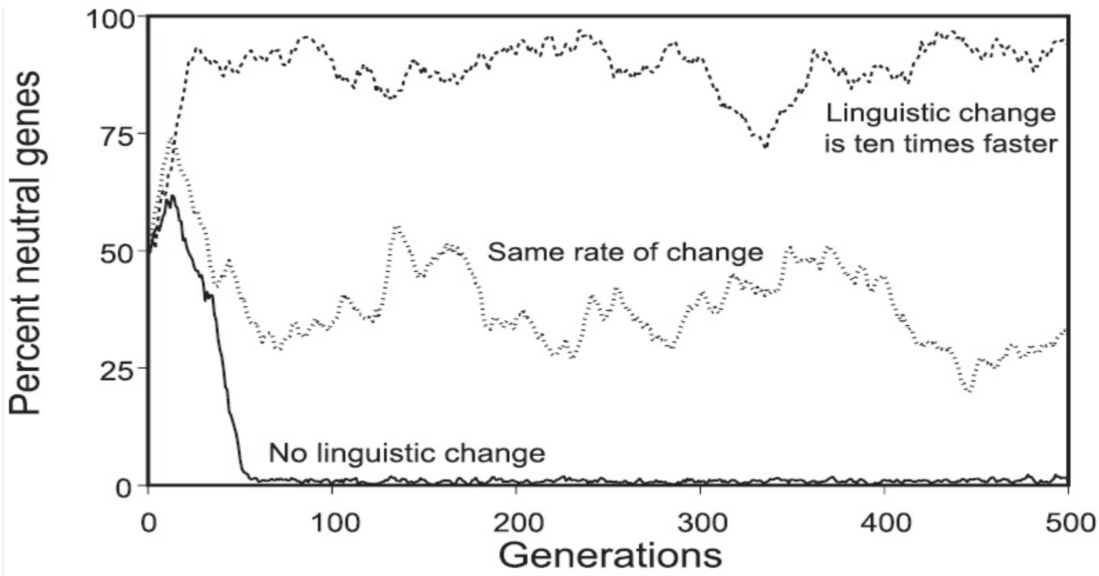


FIGURE 3.11: Reproduced with permission from [48]: Neutral genes that can learn language features slowly are plotted at each generation. A high rate of change of language prevents genetic assimilation whereas a fixed linguistic target enables assimilation via the Baldwin Effect.

Although there is an assumption in the results in Figure 3.11 that non-neutral genes are fixing on the correct language target trait values, it is safe to assume that this simple experiment demonstrates that a high rate of change of gradient to the adaptive peak prevents assimilation.

3.4.2 Social Learning and Learning Stability

Social transmission of learning is often thought to be an effective exo-genetic inheritance mechanism that facilitates inter-generational transmission of learning. Whilst social learning could be viewed as removing the need for the Baldwin Effect, if social learning is costly then there is still an evolutionary benefit for the behaviour to become innate. Again, the predator example being an extreme case - waiting to see how the rest of a herd avoids a predator may be a foolhardy strategy. Social learning could be a mechanism by which the consistency of learning across generations facilitates the Baldwin Effect with the learnt outcome providing a fixed evolutionary target. However, the effectiveness of social learning has been challenged in a review of experimental and theoretical work on the social transmission of learning by Claidiere et al. [19]. This work suggests that whilst imitation and other forms of social learning may be effective at spreading a learnt behaviour across the population, the transmitted behaviour quickly tends towards the propensities of the individual. This degradation is influenced by factors such as the number of naïve learners, the time available for individuals for exploration and the number of social cues [19]. This would suggest that social learning alone may not provide the consistency of learning outcomes across large numbers of generations that are likely required for genetic assimilation.

3.4.3 Trait Complexes and Recurrence of Learning

As discussed in Section 2.7.3, West-Eberhard [129] suggests that the selection pressures that learning can place on physical traits can make that learning more likely to occur in subsequent generations. This, according to West-Eberhard could bind the behaviours and physical traits together to become a unit of selection. An example of this is the observation that when a new species of ground-dwelling predator lizard (*Leiocephalus carinatus*) was introduced to an island, the species of lizard it preyed upon (*Anolis sagrei*) initially evolved long legs which facilitated running away from the predator but, after only six months, the behaviour changed to climbing trees and the selection pressure moved in the opposite direction, favouring shorter limbs [39]. Connected to this idea is Heyes et al.s' [48] concept that fitness improving learning causes a selection pressure on sensory organs for improvement of information fidelity. These effects could be an important mechanism for learning to be stable across the generations and therefore facilitate the assimilation of that learning. However, that would also suggest that the environment also needs to be consistent across generations for the learning physical-trait complex to remain adaptive: a significant change in environmental conditions could break the trait-behaviour complex making it non-adaptive - the speed of genetic change is the limiting factor.

Crucial to this thesis, is the idea that genetically correlations between innate behaviours can constrain learning in a way that makes the recurrence of that learning more and more likely. This recurrent learning then reinforces the evolution of these genetically specified traits and so there is mutual feedback that gradually assimilates the learning.

3.4.4 Reliability of Cue

Another way in which learning can be inconsistent, potentially inhibiting the assimilation of that learning is if the learning is of poor quality, either because the mechanism of learning is immature or because the cues to the learning are unreliable. In this scenario, the same environmental input could result in different fitnesses for the phenotype after the same learning, which has the potential to create noise in feedback from learning to evolution. The empirical and theoretical studies of cue reliability and learning focus on the evolution of plasticity itself rather than how reliability may affect assimilation. For example, in their review of Bayesian models of development, Stamps & Frankenhuis [111] identified that Bayesian models of development support earlier studies that show cue reliability can limit plasticity without a specific cost to plasticity. This would seem to be intuitive: if the cue is unreliable then the fitness benefit of the plasticity is likely to be reduced. In addition, Bayesian models suggest that the confidence of a naïve prior (in the form of genetic inheritance, parental effects and epigenetic influences) may limit the development of plasticity: the stronger the inherited factors, the less likely environmental cues are to change the phenotype.

This suggests that where cue reliability is poor, it is unlikely that the plasticity itself would have evolved and therefore there is no scope for the plasticity to eventually become assimilated. On that basis, we assume that the reliability of environmental cues to learning is a necessary condition for the evolution of learning which negates consideration of it as a condition for the genetic assimilation of learning.

3.5 Discussion

The exploration of the conditions for genetic assimilation of learning in this chapter suggests that for traditional models of plasticity that consider plasticity at an individual trait level:

1. Whilst a cost of learning does create a selection pressure for the genotype to move towards the learnt phenotype, this is not a universally necessary condition. Whilst it seems unlikely that there would be no incremental cost of learning over innate behaviours, if were the case, genetic drift can be a powerful force in

genetic assimilation of plasticity - even when that assimilation is across multiple traits. Further, if learning moves to a local optimum, then a cost of learning may anchor the genotype to that optimum, thereby inhibiting travel to the global optimum.

2. The concept of neighbourhood correlation is, as Mayley suggests, dependent on the relative distance of movement in genotype space and phenotype space, where the relative fitness effects of an individual learning step as compared to mutation dictate the probability of assimilation. However, in Mayley's specific model, lack of assimilation due to uncorrelated movement is also dependent on the search of the phenotype space being restricted; if learning changes at each generation then it will find a combination of learning where the genotype can assimilate to the learnt phenotype.

3. There needs to be consistency to learning across multiple generations for learnt behaviours to become innate. The constraints introduced by physical traits or innate behaviours may help consistent learnt behaviours be discovered at each generation. Social learning, being a mechanism to transmit learning across generations, may also help learning to be consistent over evolutionary time but its fidelity may be diminished at each generation.

The conditions for learning to become assimilated are therefore not clear cut. However, as we shall see in the next section, a more holistic view of these conditions can be developed when considering them in terms of the fitness signal between variation of the genotype and selection.

3.5.1 Learning and the Fitness Signal from Variation to Selection

To better understand why the mechanisms underpinning genetic assimilation are constrained by certain conditions, it is worth considering how learning changes the *'fitness signal'* between variation and selection. Variation can be an adaptive mutation that provides a positive fitness signal (increased fitness) or a deleterious one that provides a negative fitness signal (reduced fitness) and both signals suitably inform selection. Neutral mutations do not send a fitness signal to selection but may build up cryptic variation in the genotype which may eventually be exposed to selection when there is a radical perturbation to the environment [68]. Without learning, selection simply acts on the fitness difference produced by mutation but there is potential for the signal to be ambiguous due to epistatic interactions with other traits. For example, under some environmental conditions, an adaptive fitness effect on one behavioural trait may be masked by maladaptive fitness effects on other traits; whereas under different environmental conditions all the fitness effects of the mutation may be adaptive. Learning has the potential to magnify the fitness signal, enabling selection to distinguish the fitness differences more clearly, providing a boost to evolutionary progress. However, even when adaptive, the action of learning also has the potential to disrupt the fitness signal through a variety of effects; the conditions for assimilation are strongly related to whether or not learning disrupts the signal between variation and selection. This interaction is summarised in Figure 3.12 and how the three potential sources of disruption to the fitness signal relate to the conditions for assimilation is discussed below.

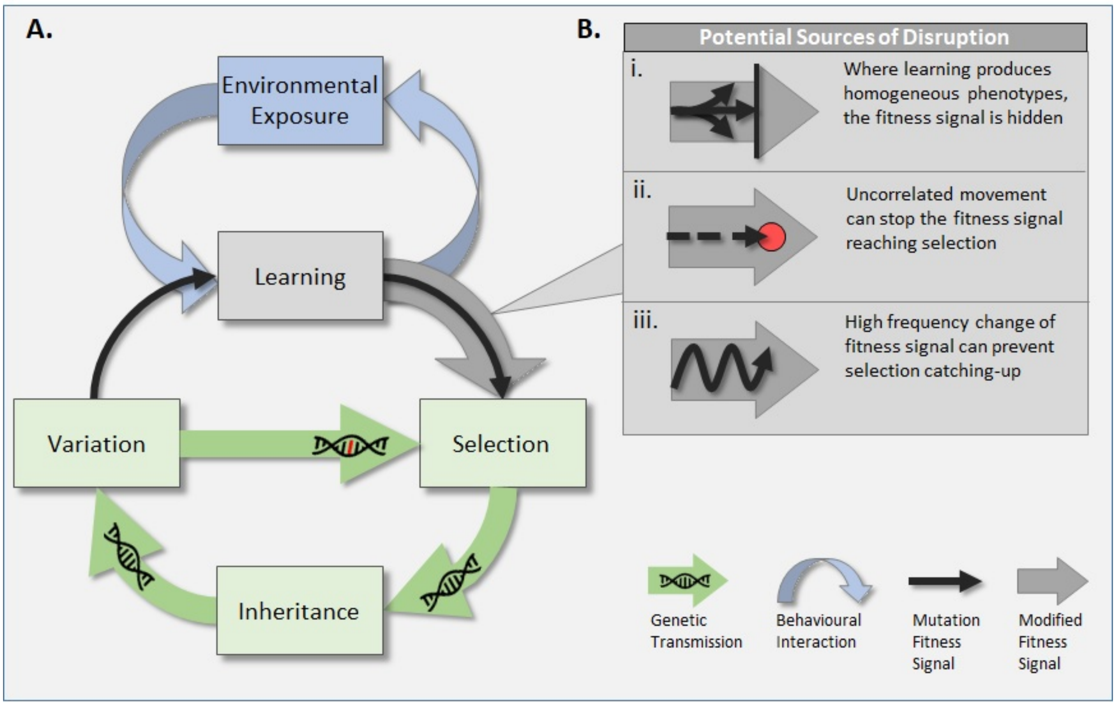


FIGURE 3.12: Learning mediates the fitness signal between variation and selection. **A.** For learning to be genetically assimilated, the fitness signal as to whether a mutation is adaptive needs to be preserved by learning. Learning has the potential to strengthen the signal (denoted by the broader arrow from learning) or disrupt that signal. **B.** Three key conditions can prevent the mutational fitness signal from reaching selection. (i) Learning may shield the fitness signal by producing the same phenotype for a wide range of neighbouring genotypes. (ii) Uncorrelated movement between genotype and phenotype may prevent the fitness signal from reaching selection. (iii) High-frequency change of learning means evolutionary change cannot effectively track with the fitness signal.

Firstly, where learning produces the same phenotype for a range of neighbouring genotypes, the fitness signal between variation and selection is effectively removed - the mutated and unmutated genotype results in the same learnt phenotype with equal fitnesses (as per the flattened landscape shown in Figure 3.1). Simple luck may move the genotype towards the learnt phenotype through drift, but there is no fitness signal to anchor the genotype to the phenotype's configuration and so full assimilation is unlikely to take place (the genotype will drift in the neutral network of the phenotype). However, if a cost of learning that is proportional to the amount of learning is applied, then the signal to selection is effectively restored; a mutation that replaces a learning step lowers the cost of learning sending a positive fitness signal to selection.

Secondly, if the fitness difference between the movement in genotype space and phenotype space is large (as is the case of Mayley's model where $L \geq 1$), a correct fitness signal cannot be consistently produced: if the learnt phenotype is close to the local optimum for a given number of learning steps, unless a learning step is removed by the movement of the genotype then the fitness signal will always be negative regardless of whether the movement in genotype space is towards or away from the local optimum.

Finally, if learning tracks a rapidly changing environment producing a variety of phenotypes, then the fitness signal can be correct (as provided by a cost of learning) but the high frequency of change in that signal makes it impossible for the genome to keep in step: slow-moving evolution cannot keep up with fast-changing learning.

From the above, one can conclude that in the broadest terms, the only condition for assimilation is that the fitness signal between variation and selection be maintained or magnified for long enough for it to have a positive effect on evolution. It, therefore, opens up the possibility for assimilation to take place in absence of the specific conditions discussed in this section. Indeed, in the models of learning and evolution presented in the next chapters, other mechanisms for maintaining the signal between variation and selection are deployed removing the requirement for a cost of learning, neighbourhood correlation and a stable learnt outcome.

Chapter 4

Fitness Landscapes and Behavioural Plasticity

4.1 Fitness Landscapes in Models of Evolution

As reviewed in Chapter 2, the fitness functions used to test evolutionary models generate either single-peaked or multi-peaked fitness landscapes[1], with multi-peaked landscapes offering the greatest challenge to the phenotype reaching a globally optimum fitness as there is potential for it to become stuck in local optimum. Consequently, as well as testing evolutionary models without plasticity, multi-peaked fitness landscapes are also a good method to test if simulations find the global optimum fitness with plasticity that are unlikely to be found without plasticity, i.e. demonstrating the Baldwin Optimizing effect as discussed in Chapter 2.

There are several methods for creating multi-peaked fitness landscapes, including NK landscapes [54], stacked Gaussian, Rastrigin, and Schwefel functions [12], and Griewank function [67], all of which all have a controllable degree of ruggedness. Whilst the stacked Gaussian, Ratrigin, Schwefel and Griewank functions - unlike the NK landscape - have predictable global optimum values, they are not generated from a building-block structure [43], where solving sub-problems leads to a local optimal fitness whilst having the right combination of sub-solutions has a globally optimal fitness. As will be argued in Section 4.2, a building-block structure may be an appropriate representation of the fitness effects of behavioural plasticity and therefore, different landscapes from

[1]Some models use a varying fitness landscape but, as discussed in Watson et al. [122], there is some degree of equivalence between moving the genotype to multiple different locations in a multi-peaked fitness landscape after a set number of evolutionary episodes and the genotype being exposed to an environment that changes after a number of generations - a Modular Varying Goals (MVG) landscape. However, Kashtan, Noor & Alon [52] identify that a varying landscape may help a phenotype reach a global optimum where fitness plateaus change into gradient to a local or global optimum.

those listed above have been chosen as test functions for examining the interaction between learning and evolution.

The problem structures that generate the fitness landscapes used for the simulations presented in this thesis - borrowed from other work [58, 125] - are constructed using building-blocks. These building-blocks are easily solvable modules but the difficulty arises in finding the correct configuration of the modules to achieve a globally optimal solution. The building-block nature of these problem types allows for structural regularities in the fitness landscape. This is important as genetic correlations can bias mutations towards fitter phenotypes [123] but usually where there are exploitable structural regularities in the landscape. Since correlations are an important structural component of the models described in this thesis, uniformly structured environments are deployed. Here we use two fitness landscapes that have significant regularities: Watson, Mill's & Watson et al.'s *'Modular Constraints (MC)'* problem [124] and Kounois et al.'s *'Concentric Squares (CS)'* [58]. In both of these landscapes, the correct combination of modules achieves one of two globally optimum fitnesses. These problems are known to be difficult to solve with simple hill-climbing algorithms as they define multi-peaked fitness landscapes where a single loci hill-climber will get trapped at a local optimum.

The MC and CC problems possess different characteristics: the MC problem encodes connections between traits that are either short-range within blocks of traits or long-range between blocks of traits, whereas the CS problem only encodes local connections. Consequently, they yield different fitness landscapes, as described in Section 4.1.1 and 4.1.2.

4.1.1 Problem Structure 1: Modular Constraints Problem

The Modular Constraints (MC) [123] problem is based on reproducing a target phenotype vector $\mathbf{t}$, of n traits defined by a set of m blocks of size k, where each block is composed of k values ($n = mk$) and where, in this example, blocks of values alternate between a positive and negative sign. This is defined as follows:

$$\mathbf{t} = (t_1, t_2 \ldots, t_n)$$

where

$$t_i = \begin{matrix} 1 & \text{if } i \bmod(2k) \geq k \\ -1 & \text{otherwise} \end{matrix}$$

As described by Watson et al.'s [123], the Modular Constraints problem is further defined by a constraints matrix C where c_{ij} defines the epistatic interaction between trait i and trait j. Each matrix element of C is set to '1' where i and j are within the same block with this value representing a strong constraint. Where the element of C matrix are in different blocks, c_{ij}, each is set to positive p where i and j in the target phenotype but have the same sign and negative p where i or j have opposite signs. Therefore c_{ij} can be defined as:

$$c_{ij} = \begin{cases} 1 & \text{if} t_i = t_j \wedge |i-j| < k \\ +p & \text{if } t_i = t_j \wedge |i-j| > k \\ -p & \text{otherwise} \end{cases}$$

In most experiments presented in this thesis, the value of p was set at 0.01 based on the relative difficulty of resolving constraints with different p values described in Watson et al. [123].

The resultant constraints matrix C is shown diagrammatically in Figure 4.1 with the strong constrained between elements within a block represented in yellow with weaker positive and negative connections between traits in different blocks represented in dark and light blue.

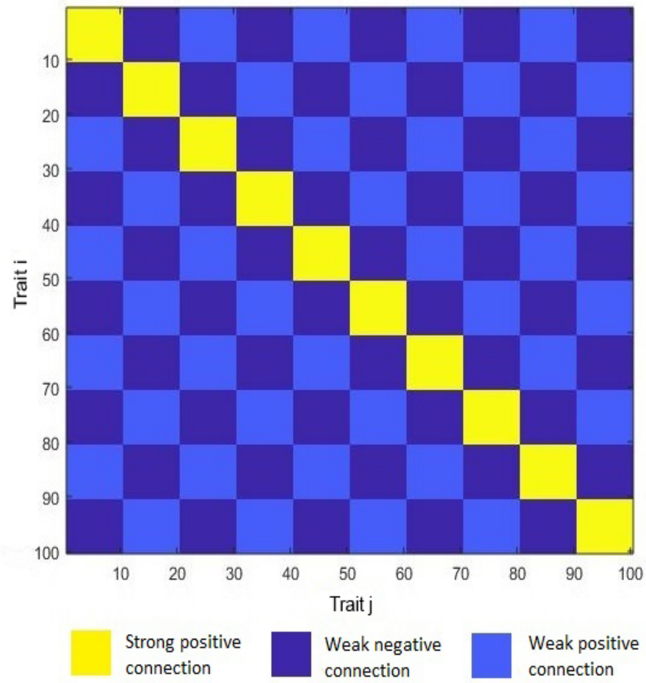


FIGURE 4.1: Illustration of constraints matrix encoding the Modular Constraints (MC) problem for ten blocks of ten traits. There are strong connections between traits within a block (represented in yellow) with weaker positive and negative connections between traits in different blocks (represented in dark and light blue)

This problem structure yields a cross-sectional fitness landscape with $m+1$ optima, where m is the number of blocks. The local optima are observed when within-block constraints are resolved, but the between block constraints are not, i.e. all traits within a block have the same sign but the signs of each block do not alternate between one block and the next as in the target phenotype. Therefore the difference in fitness value between local optima is controlled by the value of p. An example fitness landscape showing a trait by trait substitution between the target phenotype and its complement for ten blocks of size ten with a p-value of 0.01 is shown in Figure 4.2[2].

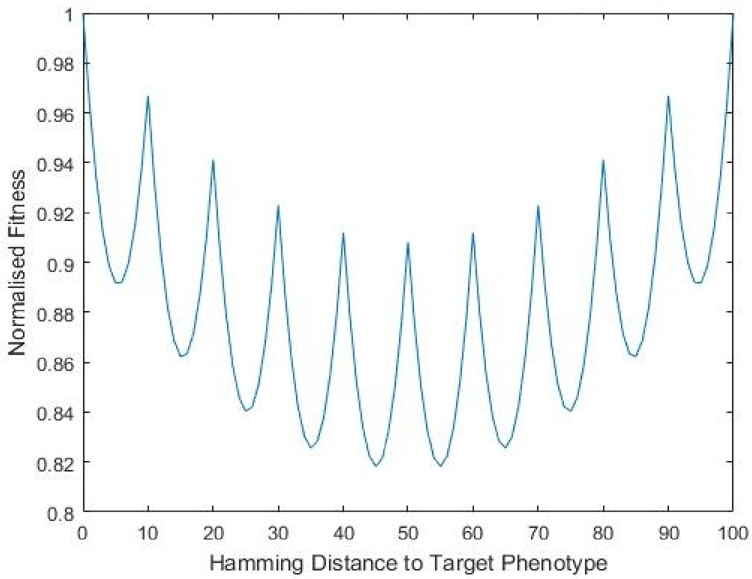


FIGURE 4.2: Example cross-section of a fitness landscape for the fully connected Modular Constraints problem. The normalised fitness of a set of phenotypes is shown as a trait by trait transition from the target phenotype (Hamming distance = 0) to its complement (Hamming distance = 100).

4.1.2 Problem Structure 2: Concentric Squares Problem

Whilst the MC problem has a clearly defined building-block structure, all traits have a specified connection to all other traits which is unlikely to be the case in the natural world. The Concentric Squares (CS) problem, borrowed from Kounios [58] has a sparser set of problem constraints with uniform constraint values representing local interactions between alleles. So that successful results can be easily identified, the constraints matrix is used to encode an image defined by a matrix M of dimensions n by n with d elements ($d = n^2$):

$$M = m_{ij} \in \{-1, 1\}^{n \times n}$$

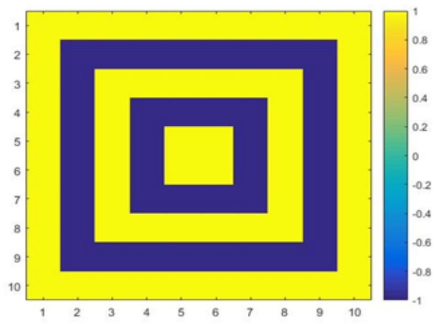
[2]This fitness landscape is generated according to the mechanism similar to that described in Kounios [58], where $\mathbf{t}$ is the globally optimal target phenotype described earlier, and $\mathbf{y}$ is set of n phenotype vectors with n traits representing a cross-section of the fitness landscape:

$$\mathbf{y^k} = (y_1^k, y_2^k \ldots, y_n^k), y_j^k \in \{-1, 1\}, j \in \{1, 2, \ldots, n\}, k \in \{1, 2, \ldots, n\}$$

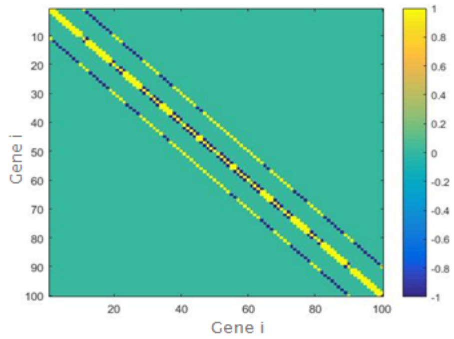
where n is, as previously defined, the number of traits in the phenotype and:

$$y_j^k = \begin{cases} t_j & \text{if } j \leq k \\ -t_j & \text{otherwise} \end{cases}$$

and where k denotes the phenotype vector with the Hamming distance from a target phenotype, transitioning to its complement where all traits are flipped.



(I) Target phenotype producing consistent epistatic constraints for the Concentric Squares problem.



(II) Epistatic network generated from encoding relationship of neighbouring pixels in target phenotype.

FIGURE 4.3: Target phenotype and resultant constraint matrix for the Concentric Squares Problem.

To build the constraints matrix, only the pair-wise sign epistatic interaction of an element of M and each of that element's four neighbours (up, down, left, right) are registered with a correlation strength of positive one where i and j are of the same sign or negative one where i and j are of opposite signs; all other interaction strengths are set to zero. In this way, the constraints can encode any image. With this problem structure, the target phenotype is defined as an image of concentric squares with the outer ring of value one as shown in Figure 4.3(I) or negative one for its complement. This makes it easy to visually identify results reaching one of the global optima. This problem is similar to finding the minimal energy state in a 2D Ising model [123].

For the convenience of calculation, this image is flattened row by row to form a target phenotype vector $\mathbf{t}$:

$$\mathbf{t} = (t_1, t_2 \ldots, t_d)$$

where

$$t_i = m_{pq}, p = \left\lceil \frac{i}{n} \right\rceil, q = i - n \left\lfloor \frac{i-1}{n} \right\rfloor$$

Using this flattened image target vector $\mathbf{t}$, each element of C can be defined as:

$$c_{ij} = \begin{cases} 1 & \text{if } i = j \\ t_i t_j & \text{if } (i = j - n \vee i = j + n) \vee ((i = j + 1 \vee i = j - 1) \wedge \lceil \frac{i}{n} \rceil = \lceil \frac{j}{n} \rceil) \\ 0 & \textit{otherwise} \end{cases}$$

This problem structure with the resultant constraints matrix is shown in Figure 4.3(II).

As Kounios et al. [58] points out, for this constraints matrix, the fitness landscape has small local optima where local blocks of constraints are solved, but resolved blocks disagree, as shown in Figure 4.4 depicting a cross-section of the fitness landscape. This problem structure does not have quite the same building-block structure as the MC problem but, in common with the MC problem, does have structural regularities within the fitness landscape. As Mills, Watson & Buckley [123] observe, exploitation of problem decomposition does not necessarily require an explicit (building-block) modular structure. Both MC and CS problems have locally optimal sub-solutions which are globally optimal if all the sub-solutions agree. Learning (and potentially evolution alone) have the potential to use the fitness signals provided by these structural regularities to change or navigate the fitness landscape to discover the globally optimal configuration. The CS problem is therefore an interesting counterpoint to the MC problem.

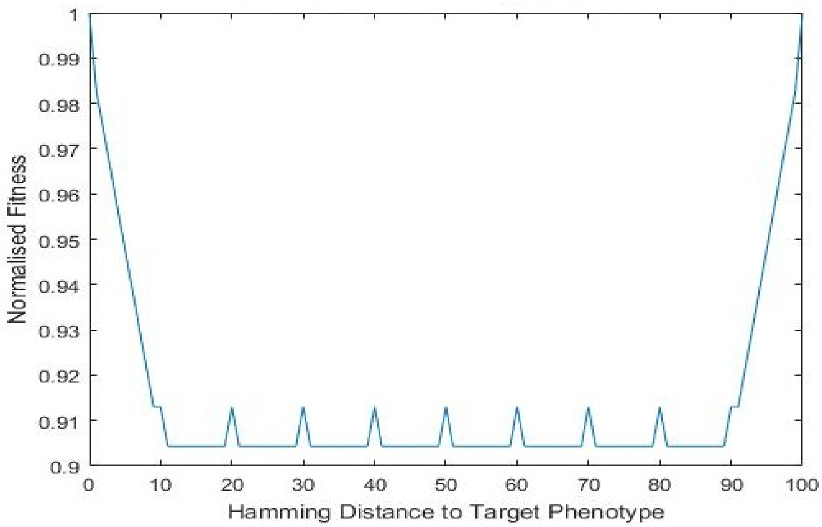


FIGURE 4.4: Example cross-section of a fitness landscape of a ten by ten Concentric Squares problem. The normalised fitness of a set of binary phenotypes is shown as a trait by trait transition from the target phenotype (Hamming distance = 0) to its complement (Hamming distance = 100). See Section 4.1.1 for detail of how this cross-section of the fitness landscape is generated.

4.2 Behavioural Plasticity and Fitness Landscapes

4.2.1 Behavioural Correlations

In the previous section, we showed how a building-block structure encoded using a constraints matrix can be used to generate a multi-peaked fitness landscape. This section considers how this building-block structure may be relevant to behavioural plasticity.

Learning is often characterised as an efficient method to search the fitness landscape allowing for spikes in fitness to be found and then be assimilated via the Baldwin Effect [129]. In the scenario posited by Heyes, Chater & Dwyer [48] this fitness spike can correspond to an individual discovering the correct combination of behaviours (their example being where a squirrel finds the right combination of *"complex, repetitive sequence of biting and hammering"* to open a nutritious nut). A similar concept is proposed by Morgan, Suchow & Griffiths [83] who use the example of bird-bathing as a *'super-trait'* that is made up of coordinated execution of multiple sub-behaviours such as pecking, raking, squatting and rolling. As they neatly put it, *"Phenotypes, whether morphological or behavioural consist of hierarchical arrangements of modular sub-units"*. Building on this theme, West-Eberhard [129] (p.59) - using the example of the courtship of male grasshoppers - also discusses how an individual's overall behaviour may be viewed as a repertoire of sub-behaviours each of which is sensitive to prevailing conditions. In the case of grasshopper courtship, the conditional behavioural response is triggered by the female's activity. This behavioural repertoire provides flexibility in the same way that morphological flexibility can facilitate a wide range of behaviours [129]. Further, Lorenz [69] suggests that behaviours are hierarchical, with a high-level behaviour requiring a degree of coordination between lower-level behaviours and Raine et al, [98] suggest that for bees: *"It is likely that behavioural traits are polygenic and linked through pleiotropies, that is correlated characters ..."*.

Together, this suggests a scenario where there are repertoires of behaviours with good combinations of behaviours having high fitness (e.g. the activities needed for effective hammering) and specific combinations that have an optimum fitness (the squirrel's best combinations of biting and hammering). Based on this, one can envisage a fitness landscape where the combination of behaviours required to reach a locally optimal fitness is built from a set of strongly correlated single behaviours to form behaviour building-blocks and the right combination of building-blocks representing the optimal behavioural phenotype. These hierarchical sets of behaviours may still be completely innate; for example, wasp offspring raising behaviours, such as nest building and family feeding, whilst being complex are still innate [70].

4.2.2 A Framework for Modelling Behavioural Correlations

To facilitate consideration of the genetic assimilation of learnt behaviours, the set of behaviours that a phenotype expresses would need to be a combination of both learnt and innate behaviours. These repertoires are environmentally sensitive and so different environmental inputs cause different combinations of behavioural expression. To illustrate this concept further, we use foraging behaviour in bees as a highly idealised exemplar of a set of correlated behaviours combining acquired and instinctual behaviour.

Many bee behaviours have a strong genetic basis and are therefore heritable, including pollen vs nectar selection, resource switching, load-size selection, round-trip time decision and foraging distance choice [88]. Bees also can learn and therefore their behavioural traits have both a learnt and innate component. For example, bumblebees have an innate ability for sonication to remove pollen, but the characteristics of this ability can be fine-tuned through learning [82]. The extensive literature studying bees demonstrates that their behaviour (learnt and innate) is complex and varies considerably between subspecies [20,51,82,89,91,98]. Consequently, it would likely be fruitless to attempt to produce a faithful replication of bee behaviour but instead we use a simplified framework of the foraging behaviour of a generic species of bee to make the case for a building-block structure for behaviours and suggest the relevance of the MC problem. The structure of this framework showing an interaction between environmental triggers and correlated behaviours is shown in Figure 4.5.

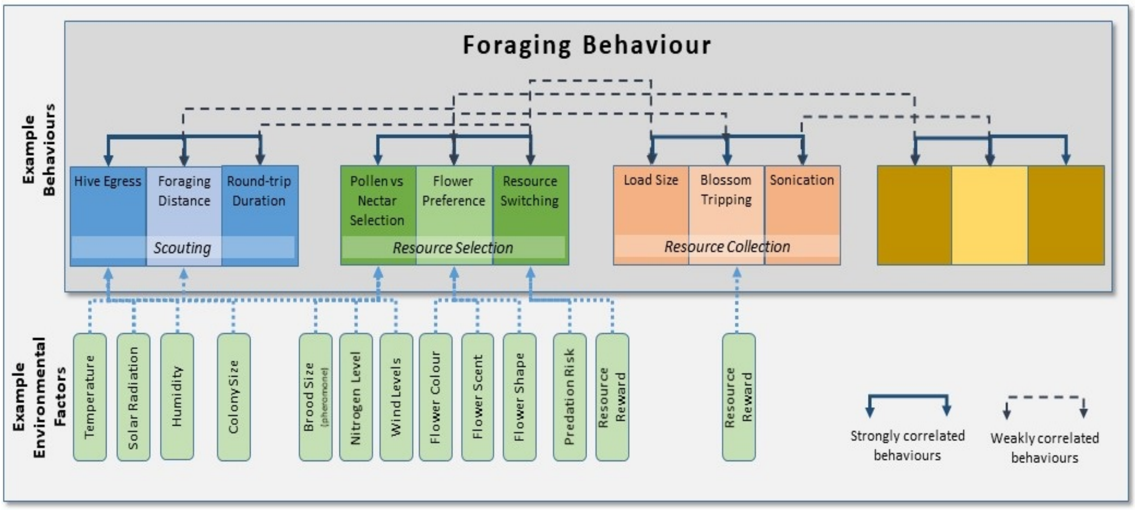


FIGURE 4.5: In this highly idealised model of a hierarchy of behaviours, sets of strongly correlated basic behaviours combine to provide a new capability, e.g. ability to scout, with a fitness benefit. Sets of behaviours can be coordinated to provide an additional fitness benefit, e.g. access to combinations of scouting, resource selection and resource collection.

Foraging for a food source requires behaviours that control the bee's actions on a scouting trip; however, to be successful, these actions need to be coordinated with other behaviours such as *blossom tripping*. In this framework, some behaviours may be highly correlated. For example, resource switching behaviours will be closely connected with flower preference as it is constrained by what flowers are available in the foraging patch. There are also likely to be weaker correlations between behaviours; for example, load-size is likely to be linked to foraging distance but will also be influenced by other factors. Behavioural traits can be both positively and negatively correlated, for example, one would expect a preference for nectar to be negatively correlated with sonication (a method of releasing pollen). For convenience of analysis, in this model, we assume that there are discrete sets of highly correlated behaviours and weaker constraints between these sets of correlated behaviours.

From the above, an individual in the population can exhibit sets of strongly correlated behaviours that provide locally optimal fitness but also has potential for good combinations of sets of less strongly correlated behaviours which give rise to optimal fitnesses. Using the bee foraging example, if all 'scouting' behaviours work in concert then the bee can successfully navigate to and from new foraging locations, which improves the fitness of the phenotype, whereas scouting in concert with resource selection, resource collection and other behaviours enable the exploration and exploitation of new sources of nectar and pollen (with an optimal fitness benefit). Based on this, one might see the similarities between behavioural expression and the MC and CS problem structures.

Bee behaviours are triggered by conditions either within the colony[3] or by factors external to the hive. For example, multiple external environmental factors trigger the decision to egress the hive to start a foraging trip: studies have shown that this behaviour is primarily driven by temperature and sunlight and to a lesser extent humidity [20]. However, for simplicity in this framework, we do not distinguish between types of environmental stimuli and it is assumed that - unlike the multiple environmental inputs to a single behaviour shown in Figure 4.5 - we assume there is only one environmental trigger for each behavioural trait. And, for the purposes of this framework, unlike Figure 4.5, low-level behaviours are assumed to be non-overlapping.

In addition, whilst the behaviours are likely to be either strongly or weakly linked from a fitness perspective, it does not necessarily mean that they will automatically become genetically correlated. As Sih [106] points out, the extent to which *'genetic behavioural correlations'* will be favoured by evolution should be a trade-off between the strength of selection to genetically correlate or decouple two behaviours, against the ease at which the connection can be made or suppressed.

This framework also ignores the effects of social learning: certain species of bee are well known for their ability to socially transmit acquired information through a form of dance. This ability is not present in all species and, as stated before, the work in this thesis does not consider the effect of social learning.

4.2.3 Applying the Framework

Using the problem structures and framework described above, we can test the interaction of learning and evolution in a multi-peaked fitness landscape and reliably measure whether learning changes the speed and/or trajectory of evolution within a loosely biological scenario. To limit complexity, in the models of learning and evolution presented in this thesis, we assume that physical traits are consistent across the population and therefore the fitness differentials between phenotypes in a population are purely down to the performance of innate and learnt behaviours in a given environment. We also assume for simplicity that all correlations between behaviours are genetically determined; correlations are not learnt and that the innate behaviours within an environment are determined solely from the correlations between innate behaviours; therefore

[3] For example, the requirement for nitrogen drives pollen foraging and, although nectar provides energy, it is not influenced to any great extent by the hive's energy needs [88].

the *behavioural trait* values themselves are not evolved. In the models in this thesis, behaviours have the potential to correlate positively or negatively (or not at all), defined by a signed weight within a correlation matrix. The specifics of this representation of genetic connections are discussed more in the descriptions of each model presented.

Chapter 5

Environmentally Sensitive Developmental Plasticity Model

5.1 Introduction

The Environmentally Sensitive Developmental Plasticity model is the first of the three new models of plasticity and evolution presented in this thesis. This model focuses on the simple case of developmental plasticity associated with morphological traits.

As discussed briefly in Section 2.1.1, the expression of morphological (non-behavioural) phenotypic traits is often influenced by environmental inputs: alternative phenotypes can be expressed depending on the environment in which the development of the phenotype occurs. Normally termed developmental plasticity, the environmental sensitivity of expressed morphological traits is a broadly accepted part of evolutionary biology with well documented experimental evidence - for example, the increase in leaf size exhibited by certain plants in conditions of reduced sunlight [112]. Whilst most commonly associated with environmentally induced changes during ontological development, the phenotypic accommodation that plasticity can provide includes examples of morphological change during more stable parts of an organisms life-cycle - for example skin-tone change after exposure to the sun. It is worth noting that these non-developmental changes to physical traits, like innate behaviours, are usually reversible (see Section 2.1.1 for a discussion of the characteristics of different forms of plasticity). Here we assume that developmental plasticity is irreversible.

Developmental plasticity has a relatively simple causal route for how the influence of the environment during an individual's development can alter the fitness of the phenotype and consequently selection, as shown in Figure 2.4 (C), reproduced for convenience in Figure 5.1.

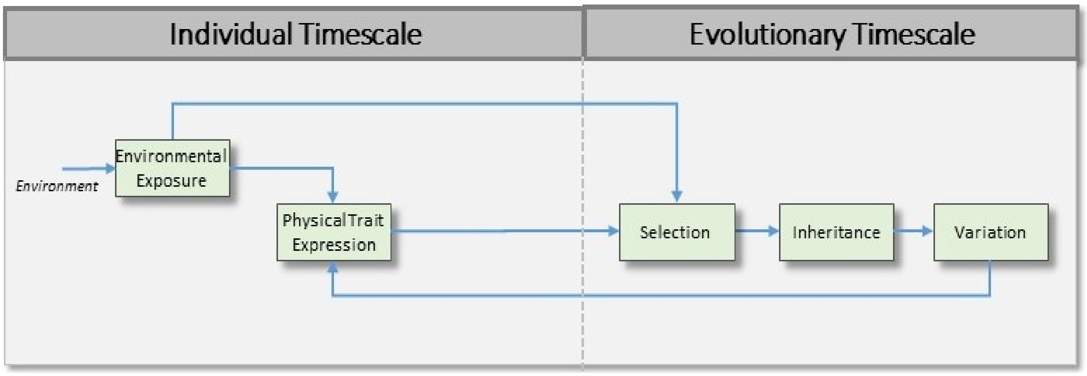


FIGURE 5.1: For developmental plasticity, the expression of physical traits is driven by environmental conditions. This morphological plasticity alters the fitness of the phenotype and so changes selection.

In Figure 5.1, the two arrows from environmental exposure, illustrate Waddington's [120] distinction between the *'epigenetic environment'* shaping the organism during development and the *'selective'* environment and also highlights that in many evolutionary scenarios, the developmental environment may be very different from the selection environment as they may be experienced in different parts of a life-cycle. Indeed, an individual can encounter many different developmental and selective environments as it progresses through several stages of a life-cycle (e.g. an embryo experiences a radically different environment to that of a neonate and also to that of the adult).

As Figure 5.1 shows there are two interacting timescales at work. Plastic expression on the individual timescale is sensitive to environmental input which produces a phenotype that becomes subject to selection over the evolutionary timescale. The causal route in Figure 5.1 is encapsulated in prior models of developmental plasticity with the effect of the environmental input usually represented as a norm-of-reaction additive to a genetically determined 'breeding value' where the plastic response is defined by a linear or Gaussian relationship to the environmental input (examples include Ancel [3,4], Lande [63], Chevin et al [17], Frank [36] and Rago et al [97]). The span of plasticity represents the environmental sensitivity of each trait independently and therefore these prior models are firmly operating in *trait space* (plasticity or evolution change trait values).

In contrast, the Environmentally Sensitive Development Plasticity (ESDP) model presented in this chapter is a significant departure from the traditional norm-of-reaction based models. It considers a scenario where the expression of traits is determined by phenotypic development based on the correlation strengths between genes and environmental inputs. This model, therefore, has much in common and shares a heritage with those developed by Siegal & Bergman [105] and Masel [72] which were based on an original model by Wagner [121]. The Siegal & Bergman and Masel models briefly surveyed in Appendix A demonstrated that canalisation can occur under a variety of conditions - for example, it can occur where there is selection for developmental stability or where there is no selective pressure. The ESDP model also has commonalities with a similar model that considered plasticity in gene regulatory circuits; Espinosa-Soto et al. [33] demonstrated that new, adaptive phenotypes resulting from a non-genetic perturbation could be stabilised through selection to become the native phenotype (also surveyed in Appendix A). Those models, therefore, focused on the reduction in plasticity; the third phase of Ancel's framework for the Baldwin Effect [4] as described in Table 2.1.

Together, these prior models suggest that if adaptive plasticity can find an adaptive peak, then pleiotropic effects can canalise that plasticity. What is needed in addition to the prior models of canalisation for a model of the Baldwin Effect[1] is a mechanism for the phenotype to find the adaptive peak[2]. To achieve this the adaptive plastic response is defined by a deterministic direct application of the environment (see Section 5.2 for detail.).

Importantly, the increased epistatic interactions also have the potential to smooth the fitness landscape [123] in a way that increases the accessibility of previously inaccessible peaks, as shown in Figure 5.2.

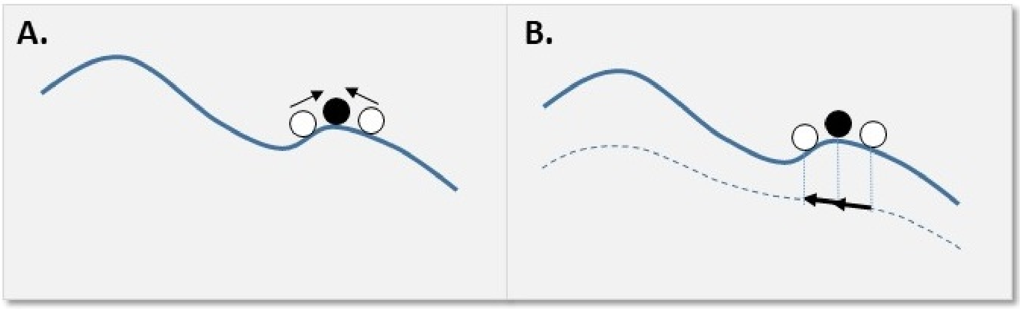


FIGURE 5.2: Adapted from Watson, Mills & Buckley [123]. (A) A phenotype (black dot) is stuck on an adaptive peak because it has a higher fitness (denoted by black arrows) than its single mutation variants (white dots). (B) Canalisation smooths the fitness landscape in a way that single mutations when measured on that landscape (dotted line) possess higher fitness and so enable phenotype to move towards the global optimum (denoted by black arrows).

Consequently, not only has canalisation a role to play in the genetic assimilation of plasticity once an adaptive peak is found but also can facilitate the discovery of higher fitness peaks. This suggests a rich dynamic between plasticity and evolution, where plasticity, creates the conditions for the discovery of high-fitness phenotypes which produces a better fitness signal between variation and selection, accelerating evolution

[1] As noted previously, the Baldwin Effect can be applied beyond purely the case of learning.

[2] It is worth noting that the Draghi & Whitlock [31] model, reviewed in Section 2.7.2, has an environmental input, uses epistatic interactions within a development loop and selects towards an adaptive peak and therefore is a gene regulatory model of plasticity. However, there several key differences between the Draghi & Whitlock model and the ESDP model presented here - the Draghi & Whitlock was not built as a model to explore the Baldwin Effect.

and potentially enabling the discovery of high fitness phenotypes that could not be found by evolution alone.

The ESDP model seeks to show that - in common with other models of plasticity - the plastic expression of phenotypes can accelerate evolution and that the plastic expression becomes genetically prescribed over time. Further, in a model where genetically specified correlations between traits evolve - a *correlation model* - evolution can be accelerated and the plastic response can be assimilated under conditions that would normally restrict the assimilation: unlike most trait-based models, this model does not require a cost of plasticity (as discussed in Section 3.2) or a stable environment creating a fixed phenotype to which to assimilate.

Developmental plasticity has different characteristics to the behavioural plasticity exhibited by innate behaviours and learning, in terms of reversibility, adaptiveness and degree to which it is genetically determined (see Section 2.1 for an extensive review). Since the claims of this thesis focus on the interaction between learning and evolution, a model of developmental plasticity may be viewed as not supportive of the core arguments. However, as well as being a novel model of the Baldwin Effect through canalisation due to epistatic effects, this model introduces some of the key concepts and structures utilised in models of the interaction of behaviours and evolution presented in Chapters 6 and 7 and acts as a step towards the more complex models of how innate behaviours and learning can affect evolutionary outcomes. Consequently, this chapter is focused primarily on exploring the dynamics of the model, rather than considering implications for developmental plasticity. However, this ESDP model does show some interesting results in its own right. Critically, it demonstrates that where an adaptive plastic response is modelled using correlations with a development function, the plastic response can find an adaptive peak and then that plasticity can be assimilated through canalisation. It also demonstrates that in common with reaction-norm models of plasticity, the action of developmental plasticity in a correlation model can significantly accelerate evolution and find global optima more reliably.

5.2 Model Structure

Although not themselves models of plasticity[3], structures from Watson et al. [126] and Kounios et al [58] are used as a basis for this model of developmental plasticity. Much of the power of those models lies in the consideration of a *correlation space* that allows a relationship between the traits that is neither mutually exclusive nor wholly dependent

[3] The Watson et al. [126] model deploy the concept of genetic correlations to demonstrate equivalence between the processes of evolution and *'Hebbian'* Learning

and with a many-to-many map between alleles and phenotypic trait values. This concept of mutually interacting structures is partially re-purposed for the computational model of the interaction between the environment and the developing phenotype.

In common with the Espinosa et al. [33] model, inputs to the model represent the environmental conditions that vary each generation[4] and there is no explicit representation of the direct effect of traits - i.e. the gene loci that are correlated are themselves not evolved. In this case, viewing the start condition as a random (adaptive and non-adaptive) environmental input maintains consistency with the behavioural plasticity models in this thesis - these models rely on environmental input as a start condition for plastic trait expression - behavioural plasticity requires environmental triggers to activate specific behaviours.

Core to the ESDP model is the formation of the phenotype that is a product of two interlocking stages: an evolutionary stage in the genotype is evolved over a set number of generations and, for each and every generation, a development stage in which the phenotypic traits are defined by an interaction between the genotype, the developmental process and the environment. Therefore, the fitness of the phenotype is determined both by the genetic disposition and the environment. Importantly, and departing somewhat from the characteristics of developmental plasticity discussed in Section 2.1.1, it is assumed that non-adaptive plastic responses to the environment will always be selected against and are therefore not explicitly modelled. This is a common assumption for models of learning and evolution with learning characterised as normally adaptive [129]. It is a less common assumption for models of developmental plasticity, where genetically specified norms-of-reaction are allowed to evolve and therefore will tend to be adaptive but can be subject to periods of maladaption, for example where there is a rapid shift in environmental conditions [97]. However, Baldwin's original theory did not consider maladaptive plasticity [22] and we assume that across a population (under the SSWM assumption) the plasticity will normally be adaptive.

The development process and environmental interaction are simplistic and assume that the plastic response will move the effective fitness of the phenotype towards a local optimum in the fitness landscape. The movement of the phenotype is achieved through a deterministic gradient ascent function, where the developing trait values are iteratively moved to higher fitnesses in the landscape for a predetermined number of development steps using the epigenetic environment. Therefore, in this model, there is only one plastic response available to an individual in the population. It is important to note the ESDP model has the ascent to an optimum phenotype and the mutation of the genetic traits both occurring in correlation space, this avoids any potential 'hiding' effects [73] (see Section 3.2.3 for discussion).

[4] Masel [72] and Siegal & Bergman [105] models have static input whereas the Watson et al. [126] and Kounios et al. [58] models rely on a sampling of inputs.

The balance of how much of a fitness contribution is due to the physical trait (via gene expression) and how much is due to plasticity must be subject to forces of selection on fitness differences (i.e. is allowed to vary over generations). In this model, this is derived from the genetic correlations encoded in the correlation matrix. As these correlations form and strengthen through selection, filtered through the fitness-enhancing effects of adaptive plasticity, they have an increasing influence on the eventual phenotype's trait values.

As discussed in Section 2.4, the characteristics of the fitness landscape can determine to what extent phenotypic plasticity accelerates or buffers the action of natural selection. In this chapter we use the Modular Constraints (MC) and Concentric Squares (CS) problem structures as introduced in Chapter 4 both of which are multi-peaked fitness landscapes with multiple local optima and two global optima; this enables the fitness of the phenotype to be assessed in a variety of environmental contexts.

Whilst not considered in the same depth as for its relevance to learning (as extensively discussed in Section 4.2), one can conceive of a scenario where the MC and CC problems could be an apt conceptualisation for correlated sets of physical traits. Under this scenario, sets of highly correlated physical traits expressed via an environmental input and a development cycle are locally optimal, and where the right combination of sets of correlated traits are globally optimal.

The 'strong selection, weak mutation' (SSWM) assumption [41] is deployed in this ESDP model, where the population can be represented by the population mean genotype and resultant phenotype. The SSWM assumption is based on the notion that a mutation would become fixed or lost in a population before the next mutation arises. Perfect elitism also is assumed, so that if mutation makes the population-mean phenotype fitter, then the mutation is always accepted.

5.3 Model Detail

The model simulating the interaction between adaptive plasticity and evolution is based on a vector representation of the population of phenotypes being the product of a matrix of genetic correlations interacting through an evolution function, a development function and a fitness function. The detailed relationship between these functions is shown in Figure 5.3 and is further defined in this section.

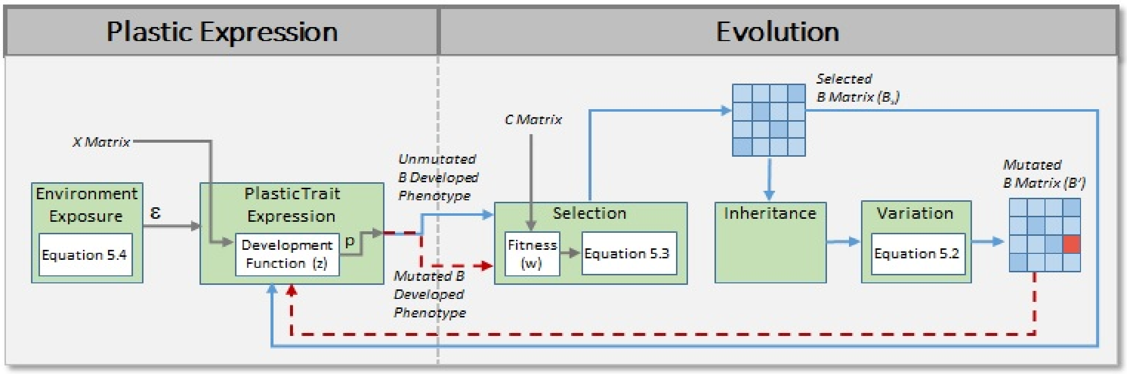


FIGURE 5.3: Following the structure shown in Figure 5.1, an illustration of the interaction between ESDP model functions. Under the strong selection, weak mutation (SSWM) assumption, the plastic expression of a phenotype for a given environment ϵ is assessed for both the unmutated (B) and mutated (B') correlations matrix using the development function (z). On the evolutionary timescale, the fitness of the resultant phenotypes (derived from the mutated and unmutated correlation matrices) is assessed by the fitness function (w) using the C matrix which varies by problem type assay (see Chapter 4). The mutation to the B matrix is generated using equation 5.2 and equation 5.3 is used to select whether the mutated or unmutated matrix becomes the next generation; inheritance is an integral part of equation 5.3. At each generation, a new, random environment is generated using equation 5.4. The flow of the unmutated form of the correlation matrix (B) and resultant developed phenotype is denoted by solid blue lines whereas the flow of the mutated correlation form of the correlation matrix (B') and derived developed phenotype is represented by red dashed lines.

5.3.1 Evolution

The population on which selection acts is encoded as a phenotype representing the population mean phenotype and is defined by a vector of n real values where:

$$\mathbf{p} = (p_1, p_2, \ldots, p_n) \in \mathbb{R}^n.$$

Since the action of plasticity on the correlation of genes is the primary concern of this model, the average genotype is simply represented by an associative network denoting the population average correlations between genes that control physical traits. In this computational model, the population mean correlation is represented in an n by n matrix of correlations B, where n denotes the number of genes in the genotype so that:

$$B \in \mathbb{R}^{n \times n}. \tag{5.1}$$

There is, therefore, the same number of traits in the phenotype as there are genes in the genotype. Each gene has the potential to correlate with every other gene and b_{ij} represents the correlation between gene i and gene j. Mutations to elements in the B matrix, therefore, represent variations in the population mean correlation.

At the start of evolution, the B matrix is initialised to zero, denoting a complete absence of correlations between traits.

$$b_{ij} = 0, \forall i, j$$

For each generation, mutations of random magnitude are applied to the correlations matrix B. So that only advantageous mutations are applied to B, a 'test' mutation in the continuous range of $\pm d$ ($\gamma \sim U(-d, +d)$) is used to generate B'. The impact of each test mutation on the phenotype following the plastic development phase is assessed using a fitness function (W); with the mutation being retained if the phenotype is fitter. The correlation element to mutate is selected at random for each of the x mutations per generation. So for each mutation step s, a random mutation matrix with one non-zero entry $J^{(kl)}$ is defined as:

$$J^{(kl)} \in \mathbb{R}^{n\times n}, k \sim U(\{1,2,\ldots,n\}), l \sim U(\{1,2,\ldots,n\}) \tag{5.2}$$

$$J^{(kl)}_{ij} = \begin{cases} 1, & \text{if i=k} \wedge \text{j=l} \\ 0, & \text{otherwise} \end{cases}$$

For each mutation step, the mutation matrix $\gamma J^{(kl)}$ is added to the B matrix, and the fitness of developed phenotype tested in comparison to the non-mutated B. Each instance of the correlation matrix $B_{(s+1)}$ is defined iteratively as:

$$B_{s+1} = \begin{cases} B_s + \gamma J^{(kl)}, & \text{if } w(z(B_s + \gamma J^{(kl)}, \boldsymbol{\epsilon} \geq w(z(B_s, \boldsymbol{\epsilon})) \\ B_s, & \text{otherwise} \end{cases} \tag{5.3}$$

In the above γ defines the magnitude of the mutation, z is the development function, w is the measure of fitness and $\boldsymbol{\epsilon}$ denotes a random start position in the fitness landscape (z, w and $\boldsymbol{\epsilon}$ are defined below). The sequence of mutations for that generation completes at $B_{(s=n^2)}$; for all experiments where the B matrix is allowed to evolve, the number of mutations of B per generation is defined as the square of the number of genes in the genotype. The magnitude of the mutation is bounded by $\pm d$, where $\gamma = U([+d, -d])$ and d has been tuned to a value that balances the expediting and optimizing effects observed. Unless otherwise stated, d is 2x10^{-5}.

5.3.2 Environmental Change

Since plasticity is not inherited, to expose the population to a range of environmental conditions, the environmental input is reset between each generation (but not during every test on the mutation of B). At the start of the plastic developmental cycle, the phenotype is set to be the same as the environment input $\boldsymbol{\epsilon}$ which is defined by a randomly generated environment vector with continuous values drawn uniformly between -1 and 1:

$$\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \ldots, \boldsymbol{\epsilon}_n), \boldsymbol{\epsilon}_i \sim U([-1, 1]) \tag{5.4}$$

5.3.3 Plastic Expression of the Phenotype (z function)

During the plastic development stage, the development function z uses the correlation matrix B to derive a set of phenotypic traits $\mathbf{p}$, where $\mathbf{p}_{t=\eta}$=$z(B,\boldsymbol{\epsilon})$ as described below.

In common with many models of development [33, 105, 126][5], at time t, the phenotype is defined as $\mathbf{p}_t$, and since development starts at a random point in the environment, as defined by the environment vector $\boldsymbol{\epsilon}$ (equation 5.4), $\mathbf{p}_{t=0}$=$\boldsymbol{\epsilon}$.

After each development step, the phenotype $\mathbf{p}_{t+1}$ is defined by:

$$\mathbf{p}_{t+1} = \mathbf{p}_t + \sigma(\mathbf{p}_t(B + X)) - \tau \mathbf{p}_t \tag{5.5}$$

Where τ is a decay rate that helps contain the overall values in the $\mathbf{p}$ vector. For all experiments in this report τ= 0.2; that which was used in the Watson et al. model [126]. The function σ is the non-linear function that is applied to all elements of the $\mathbf{p}$ vector; in this case a hyperbolic tangent: $\sigma(x) = tanh(x)$. This is important as it facilitates a multi-modal distribution of phenotypes [126] as well as capping the value of the product of plasticity and correlation weight.

After η development steps, the plastic development is complete, and the fitness of the final phenotype $\mathbf{p_c}$ is ready for assessment ($\mathbf{p}_c$=$\mathbf{p}_{t=\eta}$). In these experiments, 20 development steps ($\eta = 20$) was chosen as a value where most developed phenotypes reached a local optimum as can be seen in Figure 5.9.

In equation 5.5, matrix X is used as a convenient method to move the phenotype to a local optimum whilst still allowing the correlation matrix B to exert influence over the final signs and magnitudes during the development of the phenotype. It represents the adaptive plasticity that influences the developmental process - Waddington's epigenetic environment and is deemed to be a small factor of the selective matrix C, i.e. the environment in which reproductive selection occurs has the same structure but a larger magnitude to the environment in which the plastic development occurs. Consequently, matrix X is derived from the constraints matrix C, where C varies dependent on problem type as described in Section 4.1.1 for the Modular Constraints problem and Section 4.1.2 for the Concentric Squares problem. Here $X = \phi C$ and so ϕ controls the influence of X during each development step. The value of ϕ= 0.04 was derived through tuning

[5]Although in their model, Siegal & Bergman set the start of development as an initial random vector, they did not associate this to being an environmental input.

of the model parameters; if ϕ is too large it takes a long time for the evolved correlations to assert an influence over the phenotype's trait values, if it is too small it does not improve the fitness signal to evolution.

For experiments where there is no in-life adaptive plastic development, the development steps still occur, but the X matrix is removed from equation 5.5 so that the production of the phenotype is only guided by the genetic correlations encoded in the B matrix that is still subject to evolution. It should be noted, the production of the phenotype remains environmentally sensitive, even with the removal of the adaptive plasticity in the form of the X matrix: the environmental inputs are the start condition to the development of the phenotype.

The fitness landscapes used in this experimental set-up (described in Section 4.1) have two phenotypic configurations that match the global optima by resolving all the sign epistatic constraints, these being the target phenotype and its complement. A simple example being where there are positive epistatic constraints between all traits, then a vector of phenotypic traits containing all ones or all zeros will be at one of the global optima. The fitness of the phenotype $\mathbf{p_c}$ is therefore assessed by the phenotype's closeness to the target phenotype or its complement as specified by the constraint matrix C - this matrix being dependent on the model problem being tested, as described in Section 4.1. For the experiments in this chapter, unless otherwise stated, for the MC problem, the between-block constraints parameter p is set to 0.005[6].

5.3.4 Fitness Measurement (w function)

Adapted from Kounios et al. [58], the fitness of a phenotype vector $\mathbf{p_c}$ is assess using the selective environmental matrix C and is based on the number of traits that have the same sign as the target phenotype (i.e. the sign of the first row or column in C) and the magnitude of the trait. This is calculated from the multiplication of the phenotype $\mathbf{p_c}$, the constraints matrix C and the transpose of the phenotype as per equation 5.6.

$$w = \mathbf{p_c} C \mathbf{p_c}^T + y \tag{5.6}$$

As discussed in Section 5.3.3, the constraints matrix C varies dependent on problem type as described in Section 4.1.1 for the Modular Constraints problem and Section 4.1.2 for the Concentric Squares problem.

[6] Analysis by Mills [79] suggests that this value makes it pathologically difficult for a standard hill-climber to reliably discover a global optima where the number of loci is large

Using this method to calculate fitness means that, even when considering a binary phenotype and within-block correlations all of the same magnitude, the fitness of the phenotype does not increase linearly but instead increases quadratically with the number of correctly signed traits, as shown in Figure 5.4.

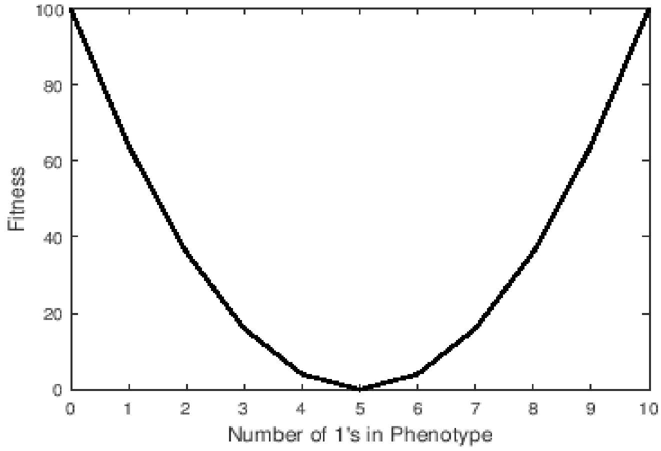


FIGURE 5.4: Cross-section of the fitness landscape for one block of ten traits.

In one of the experiments, the effect of noise in the fitness function is assayed. This reflects the biological reality that selection itself has a random component, i.e. there is an element of luck to whether a high fitness phenotype is selected. Where developmental noise is added to the fitness function, y is a random value drawn uniformly between $-\psi$ and ψ:

$$y \sim U([-\psi, \psi])$$

For experiments without noise $\psi = 0$.

Since it is the sign of trait that determines whether it matches the target rather than the magnitude, when comparing experimental results, it is often necessary to threshold the fitness function to counteract differences in fitness values due to non-comparable B magnitudes. For example, in control experiments where no evolution of the B matrix takes place, the magnitudes of the traits encoded in vector $\mathbf{p}$ are very different to an experiment with the evolution of B. The thresholded fitness function is:

$$w'(\mathbf{p_c}) = w(\theta(\mathbf{p_c})) \tag{5.7}$$

Where w is the fitness function and θ thresholds all elements of the vector $\mathbf{p_c}$ and is defined in equation 5.6.

$$\theta(x) = \begin{matrix} 1, & \text{if } x \geq 0 \\ -1, & \text{if } x < 0 \end{matrix} \tag{5.8}$$

5.4 Results

In Chapter 2, two potential Baldwin Effects were discussed: the Baldwin Expediting Effect [3] where plasticity accelerates evolution and the Baldwin Optimizing Effect [133] where plasticity can discover, and then be assimilated to, high fitness phenotypic configurations that cannot be found without plasticity.

Three basic experiments have been conducted to test if these effects are observed in this Environmentally Sensitive Development Plasticity model. The first tests developmental plasticity and evolution working together on the Modular Constraints (MC) and Concentric Square (CS) problem. The MC problem is then used to test the plastic expression of the phenotype without evolution. Thirdly, evolution without plastic expression of the phenotype is assayed for both the MC and CC problems. After those basic experiments, a further test is made to determine if plasticity makes evolution more robust against noise in the fitness function. The following sections describe each experiment and the key results.

5.4.1 Evolution with Plasticity Converges on Globally Optimal Phenotypes

The first experiment is designed to show that the combination of environmentally sensitive development plasticity combined with the evolution of genetic interactions can reliably reach one of the two global optima over the evolutionary timescale for the two problem structures. The ability to reliably reach a global optimum suggests the potential for a Baldwin Optimizing effect (as discussed in Section 2.5), where plasticity attains fitness optima that would not be attained without plasticity. To test this, a simulation is run for 10,000 generations with both the deterministic gradient ascent to a local optimum (representing adaptive developmental plasticity) and mutation of the correlations matrix (representing evolution) occurring at every generation as described in the experimental setup.

The plasticity case in Figure 5.5 confirms that, with plasticity, at the start of evolution the globally fit phenotype is not found but by the end of evolution the target phenotype and its complement, both being global optima, are reliably produced. For the same set of parameters, Figure 5.5 also shows that without plasticity a globally optimum is not reliably produced. The no plasticity case is further tested in Section 5.4.3

To verify that plasticity first produced locally optimal phenotypes and then globally optimal phenotypes, Figure 5.6 shows the phenotypes produced at the start and end of evolution for the Modular Constraints problem. For ease of comparison, the phenotypes are shown in a form where a square in each column represents one trait of the phenotype produced at the end of that generation. As can be seen in Figure 5.6,

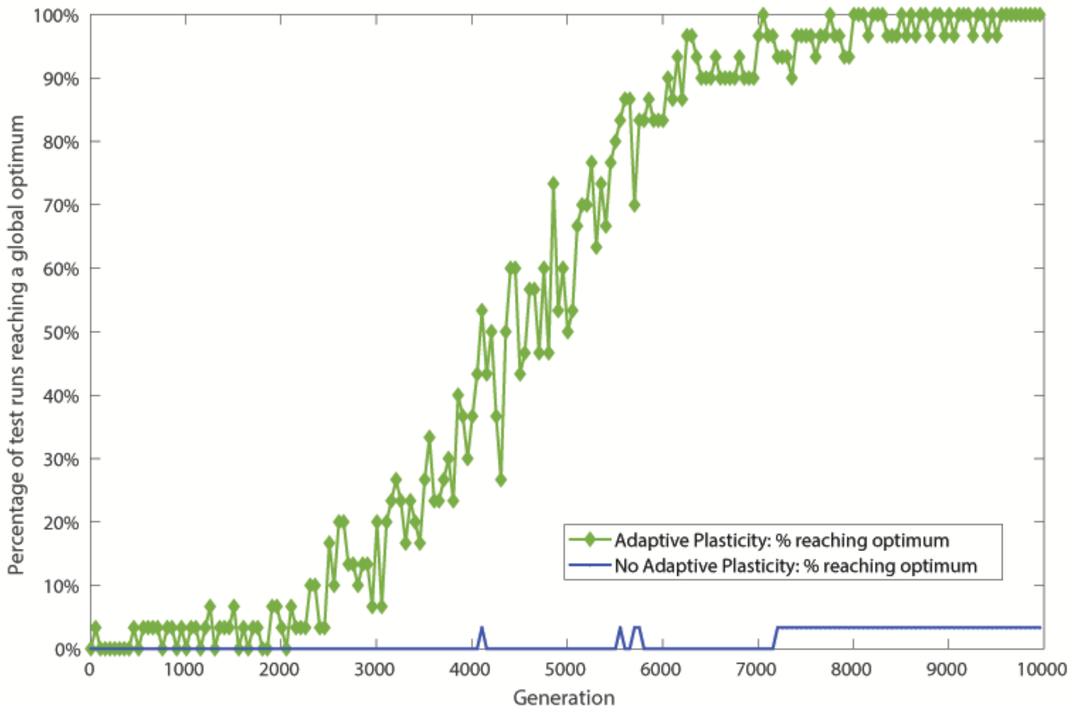


FIGURE 5.5: Comparison of evolution with adaptive plasticity and evolution without adaptive plasticity for the MC problem. With plasticity, shown in green with diamond markers, the global optimum is found for 100% of the 30 runs. Without plasticity, shown in blue (no markers) a global optimum is not reliably found but is encountered more often than it would be by chance.

in early evolutionary cycles (generations 1 to 10), phenotypes that have a locally optimal fitness are developed, i.e. the traits in each block of ten agree to satisfy the strong within-block constraints but these blocks are not in the alternating pattern showing that weaker between-block constraints are not satisfied. This confirms that the plasticity is finding locally optimal phenotypes but not globally optimal phenotypes. However, by the end of the evolutionary phase (generations 9990 to 10000), the traits in the phenotypes are showing the two alternating patterns that satisfy the weak between-block constraints. Therefore, the correlation matrix combined with developmental plasticity together reliably find the optimal phenotype regardless of the environmental input.

Next, we compare the performance between the MC and CC problems. Figure 5.7 shows a graph of the thresholded fitness of the phenotype representing the mean of the population plotted at every 10 generations of one run of the simulation. Each plot point indicates the fitness of the thresholded phenotype at a generation and after development with plasticity. Part (I) of the figure shows that for the MC problem the phenotype's fitness starts by regularly converging on locally optimal values and finds the global optima fitness with increasing frequency after each generation until by the end of evolution only optimal fitness phenotypes are discovered. Using the same set of parameters, Part (II) of Figure 5.7 shows a similar same result for the Concentric Squares (CS) problem where the combination of adaptive plasticity and evolution finds

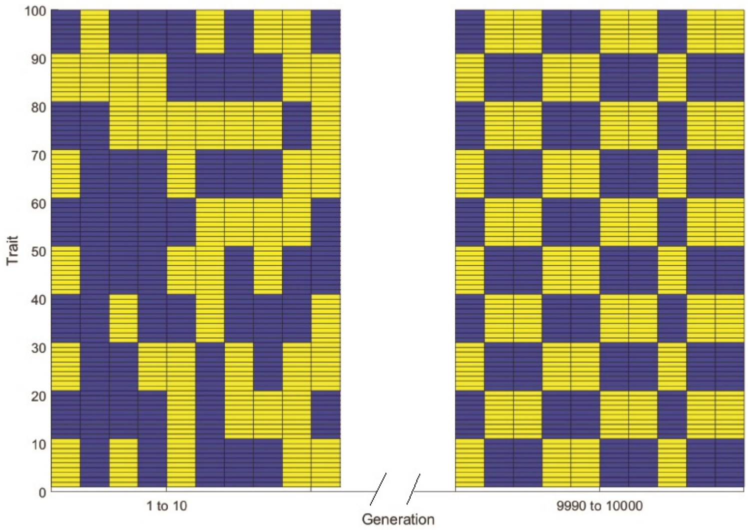


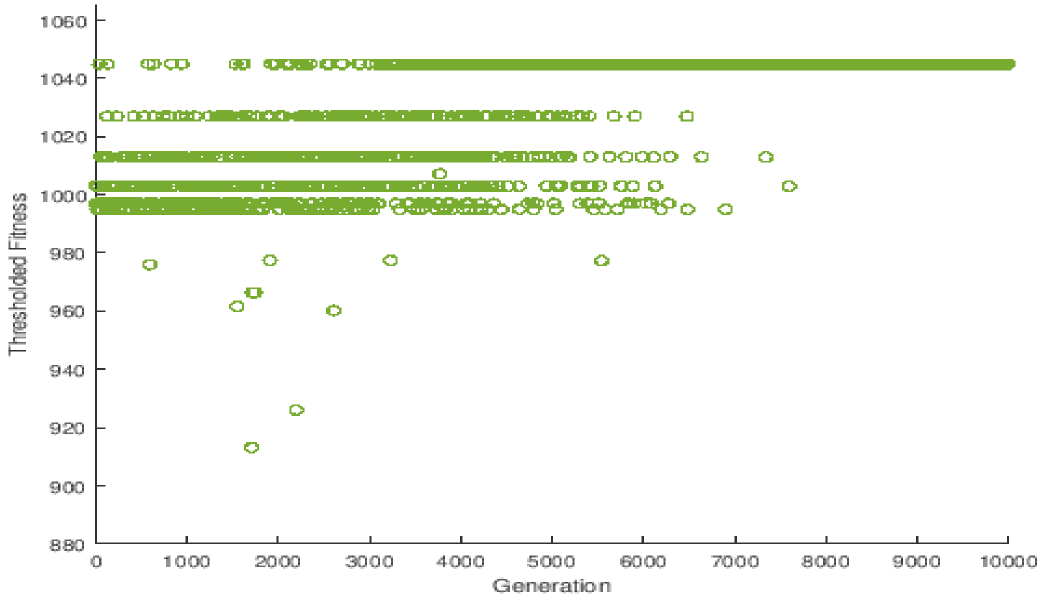
FIGURE 5.6: A sample of expressed phenotypes for the Modular Constraints problem. Each column shows a phenotype vector at a generation where yellow depicts a trait with a positive sign and blue depicts a trait with a negative sign. Phenotypes as shown for the first ten generations are consistent with being at the local optima as they are random blocks of traits of a consistent sign. The last ten generations (9990 to 10000) are consistent with being at a global optimum - alternating blocks of consistent sign.

the global optimum fitness. This result demonstrates that the action of the adaptive plasticity is effective on multiple problem structures and can resolve both the long-range constraints present in the MC problem and the local constraints characterised by the CS problem.

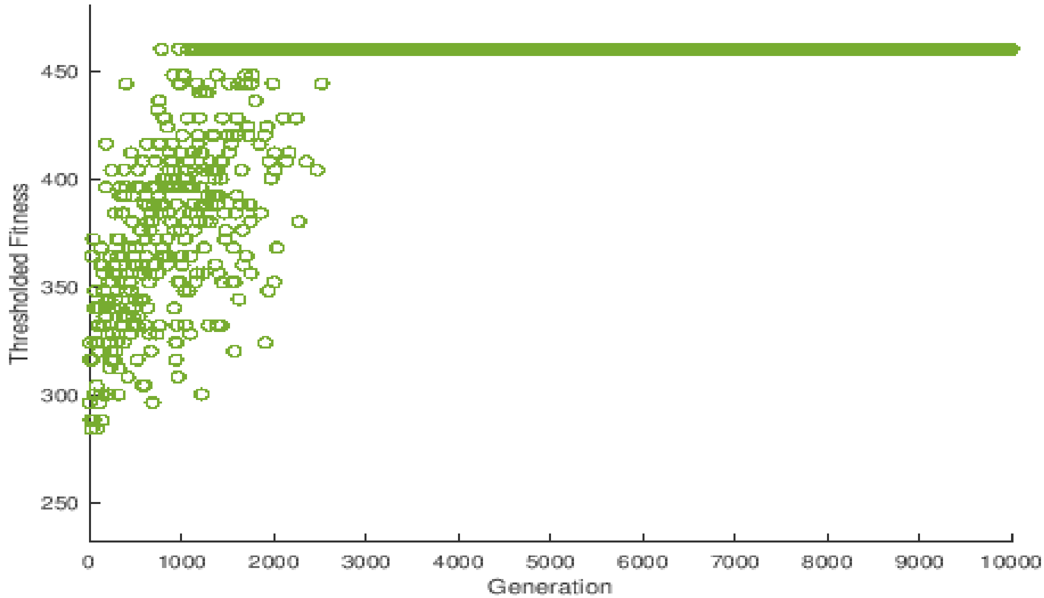
It is interesting to note that for the MC problem, the evolved correlations between genetic traits (as encoded in the B matrix) mirror that of the constraints matrix C shown in Figure 4.3, albeit with the ratio between the weight of between-block connections and weight of the within-block connections being much closer in the evolved B matrix. However, for the CS problem, the correlations between the sparsely populated constraints matrix (as shown in Figure 4.3) leads to a B matrix that has correlations between all genes as shown in Figure 5.8. If connections are possible, then correlations will evolve. All traits being able to evolve to connect to all other traits is common in models of plasticity with similar structures (for example, the models of Siegal & Bergman [105] and Masel [72]) but may be considered by some as not a biologically realistic scenario. Whilst not explored further in this chapter, limiting the number of connections that can be evolved is investigated in Chapter 6.

5.4.2 Plasticity Without Evolution is Constrained to Local Optima

This next experiment is designed to confirm that, in the absence of evolution of the genetic correlations, the plastic phenotype would not reliably discover the global fitness



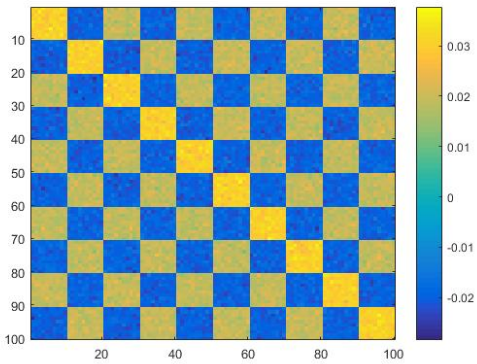
(I) Problem 1 - Modular Constraints Problem



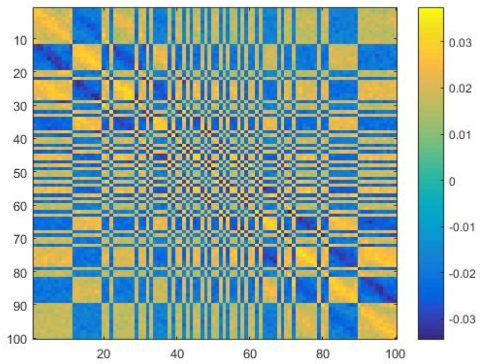
(II) Problem 2 - Concentric Squares Problem

FIGURE 5.7: (I) The fitness of the population mean phenotype after each evolutionary episode for MC problem with both plasticity and evolution of correlations. Each point on the chart represents the thresholded fitness of the developed phenotype at the end of that generation. Fitnesses can be seen to mostly converge on local optima and then converge on the global optimum. (II) The fitness of the mean phenotype for each evolutionary episode for Concentric Squares problem with both plasticity and evolution of correlations. Unlike the MC problem, convergence on local optima is not quite so clear as there is limited fitness differential between local optima and the local peaks are small compared to the global optima. However, the thresholded fitnesses do converge on the global optima.

optimum but instead would continually find local optima. To verify this, a simulation is run over 10,000 generations but where all mutation of the genetic correlation matrix



(I) MC Problem



(II) CS Problem

FIGURE 5.8: Evolved B matrices at the end of evolution for the MC and CS problems. Yellow-toned pixels indicate a positive weight in the correlation matrix whereas blue-toned pixels indicate a negative weight in the correlation matrix. For the MC problem, the evolved B matrix encompasses the constraints encoded in the C matrix. For the CS problem, more correlations are encoded than are present in the C matrix due to the fully connected nature of the B matrix. (I) Correlations in the B matrix at the end of 10,000 evolutionary episodes for the MC problem. (II) Correlations in the B matrix at the end of 10,000 evolutionary episodes for the CS problem.

is inhibited so that correlations in the B matrix do not evolve. The developmental plasticity still consistently reaches the nearest local optima through a deterministic gradient ascent.

The results of the experiment shown in Figure 5.9 confirm that the plastic phenotype regularly reaches a local optimum, that on occasion by chance, it also encounters one of the two global optima, but that the consistency of finding a global optimum does not increase over evolutionary time.

5.4.3 The Global Optimum Can Be Attained Without Plasticity

From the results above, a core hypothesis is that plasticity finds the local optima and it is the improvement in fitness signal provided by plasticity that allows the correlations to evolve so that the correlation matrix combined with the plasticity allows the phenotype to converge on a global optimum. One would, therefore, expect that turning off the plasticity function would prevent the correct correlations from being found and therefore no local or global optimum would be discovered. If this were the case it would serve as confirmation of a Baldwin Optimizing Effect where the plasticity is enabling the discovery of fitness optima that cannot be found without plasticity.

The no adaptive plasticity case shown in blue in 5.5 appears to support this intuition: evolution with random environmental inputs is discovering a globally fit phenotype

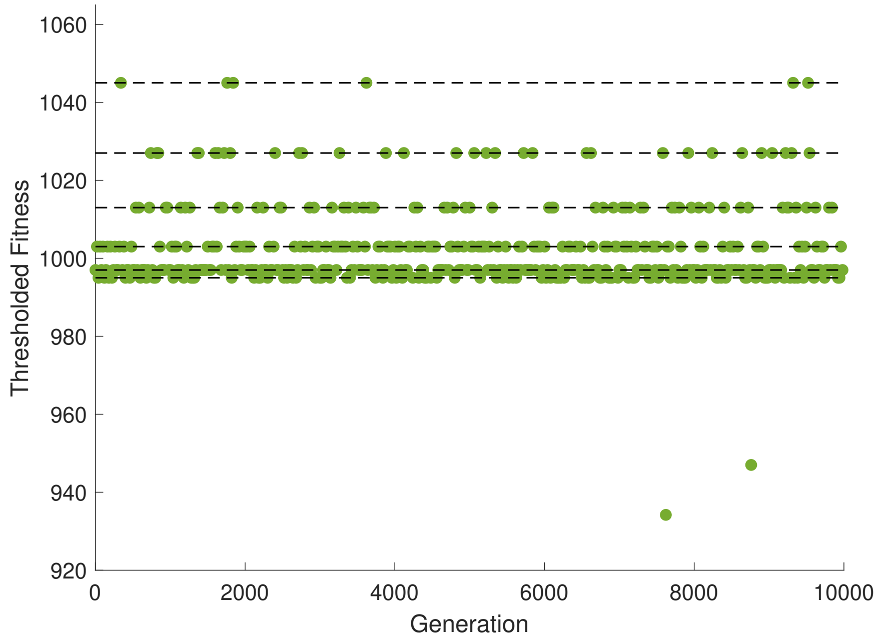


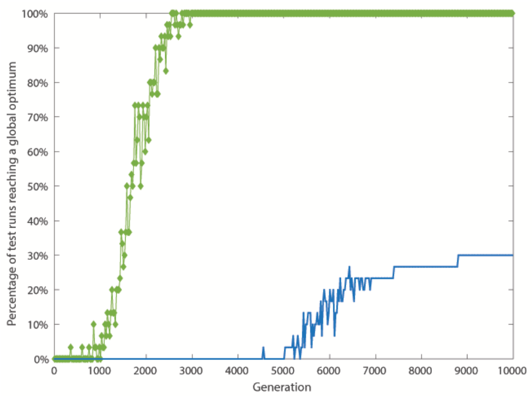
FIGURE 5.9: Comparison of phenotype fitness at the end of each generation to the peaks in the fitness landscape representing locally optimal fitnesses. Each dot shows the thresholded fitness of the population-mean phenotype at the end of each generation. Without evolution, the distribution of fitnesses does not change between generations. The black dashed lines show the locally optimal fitness values, the highest one being the globally optimum fitness. The plastic expression of the phenotype normally reaches one of the local optima, but there are a few instances in which a local optimum is not reached: those not positioned on a locally optimal fitness line.

on occasion by chance. However, as Figure 5.5 shows, after generation 7,000, the global optimum phenotype is found more consistently than would be found by chance[7].

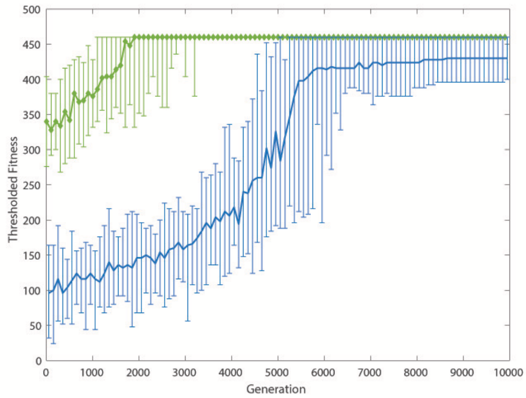
Using the same parameters as for the MC assay, this effect is shown to be exaggerated in the CS problem as depicted in Figure 5.10, one of the two global optima are consistently found in about 30% of the runs without plasticity as compared to 100% of the runs with plasticity, albeit it takes longer for the optimum phenotypes to be discovered without plasticity. A similar result has been obtained for the Modular Constraints problem; where slightly stronger between-block constraints in the problem matrix C are used ($p = 0.01$). This seems a perplexing result as with this p value local optima are still present and hill-climber would not normally find one of the global optima (as per analysis performed by Mills [79]). In addition, the environmental inputs are random, and consequently, there should be no fitness signal for selection to act upon: mutations to the correlation matrix are just as likely to be correct as incorrect. It is therefore difficult to understand how a globally optimal phenotype is generated so frequently. Much

[7] If $p = 0$ then all basins of attraction in the fitness landscape will be of the same size. Consequently, for ten blocks of size ten, the probability of finding one of the two global optimum phenotypes is $\frac{2}{2^{10}}$ which is 0.20%. However, as p grows, the basin of attraction for the global optima enlarges. Using a stochastic restart hill-climber to find local optima shows the global optima would be encountered 0.34%, 0.23% and 0.20% for the p values of 0.01, 0.005, and 0.001 respectively.

time has been dedicated to trying to understand this effect and whilst not conclusively shown for the whole system dynamic (i.e. the build-up of mutations to correlations over evolutionary time is not considered), Section 5.5.3 explains why there is a signal from random environmental vectors when considering correlation-based models.



(I) Percentage runs reaching a globally optimal fitness with and without adaptive plasticity.



(II) Fitnesses achieved at each generation with and without adaptive plasticity.

FIGURE 5.10: Comparison of fitness of the plastic phenotype versus non-plastic phenotype for the CS problem. (I) Percentage of 30 runs reaching the globally optimal fitness with and without plasticity. (II) Fitness for each generation is plotted for the same assay. The error bars show the min to max range over 30 runs using the same parameters, whilst the darker line shows the median value over the same 30 runs. With developmental plasticity, shown in green with diamond markers, the maximum fitness phenotype is quickly found. However, in this case, the locally optimum phenotypes are consistently found without plasticity (shown in blue with no markers). Data points are shown at every 50 generations.

5.4.4 Plasticity Insulates Evolution from Noise

To help verify that the signal to evolution in the no plasticity case in Figure 5.10 is weak and also determine if the strong signal from plasticity helps evolution be more robust, the MC problem was tested by adding a small amount of noise to the fitness function. This was achieved by using the fitness as per equation 5.6 with the noise parameter ψ set to 2×10^{-6}. The magnitude of the noise was determined so that it is greater than the magnitude difference caused by a mutation to the on-block at the start of evolution[8]. The fitness difference of the average mutation in this scenario being 1.3×10^{-6}. This level of noise will therefore disrupt the fitness signal of a mutation at the start of evolution.

As can be seen from Figure 5.11, the noise applied to the fitness function masks the small fitness signal that selection receives from the random phenotypes and therefore without plasticity the phenotypes do not materially improve their fitness. Where the

[8]Mutation of the within-block trait correlation provides the largest fitness difference - for a blank B (at the start of evolution) and an average environmental input magnitude of 0.5.

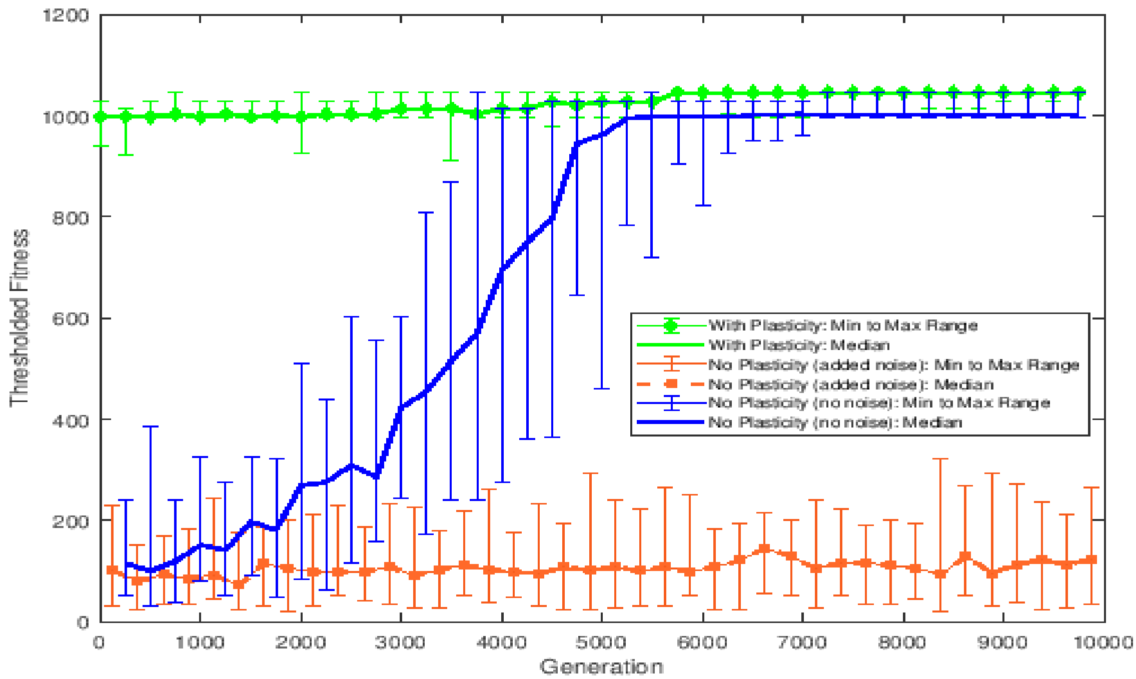


FIGURE 5.11: Comparison of evolution with plasticity and evolution without plasticity in a noisy environment. Dynamic noise is added to the fitness function. The error bars show the min to max range over 30 runs using the same parameters, whilst the darker line shows the median value over the same 30 runs. With plasticity and with noise, shown in green with diamond markers, the global optimum is found. Without plasticity and with noise, shown in orange with square markers the fitness does not increase. For comparison, the blue line without markers shows the no learning case in the absence of noise. Data is plotted every 250 generations.

plasticity moves the phenotypic fitness to a local optimum, the fitness differences between a mutation of B and no mutation of B will be larger and therefore the noise in the fitness function will not drown out the signal to selection. In this way, the plasticity insulates evolution against noise allowing evolution and plasticity together to discover globally optimal phenotypes.

This result highlights that at the beginning of evolution, the development function causes the resultant fitness difference between the mutated and unmutated phenotype to be smaller than the size of the mutation to the interaction. Consequently, the ESDP is very sensitive to relatively small amounts of noise. This suggests that correlation-based models that use a development function (for example, Watson et al.'s developmental memory model [126]) are also likely to be relatively fragile to noise.

However, this experiment also clearly shows that the increased fitness magnification effect provided by plasticity makes evolution more robust against noise in the fitness function.

5.5 Analysis and Discussion

5.5.1 Evidence for a Baldwin Optimizing Effect and a Baldwin Expediting Effect

One objective for the construction of the ESDP model was to demonstrate the Baldwin Optimizing Effect. Although evolution together with plasticity reaches fitness optima reliably that would be pathologically difficult for a simply hill-climber, with the right parameters so will evolution alone. The reason that the no plasticity case can evolve the correct correlations is subtle (as will be set out in Section 5.5.3) and the relatively weak signal that is present in the no plasticity case can be removed with the application of a small amount of noise. So whilst a Baldwin Optimizing effect cannot be claimed under all conditions, it can be claimed that the ESDP model demonstrates this effect under a limited set of conditions.

However, where both the plasticity and no plasticity case express the globally optimal phenotype Baldwin Expediting effect is clearly observed: the plasticity accelerates evolution.

5.5.2 The Impact of Plasticity and Evolution on the Fitness Landscape

To understand the action of plasticity and evolution in this model, it is worthwhile examining the effect that each of plasticity and evolution individually has on the fitness landscape.

As discussed in Section 2.4, in prior trait-based models of plasticity, the action of that plasticity during development works to flatten the fitness landscape. For example, the fitness landscape in Borenstein et al.'s [12] mathematical analysis of deterministic ascent in a multi-peaked fitness landscape (modelling learning) is a series of plateaus as shown in Figure 3.1. As shown in Figure 5.12, in this ESDP model a similar effect of plasticity is observed.

However, the evolution of the correlation matrix also changes the fitness landscape in a way that is qualitatively different from the action of plasticity. Figure 5.13 shows that as the correlation matrix evolves, it alters the shape of the fitness landscape gradually making it smoother, thereby removing local optima. This aligns with Watson, Mills & Buckley's assertion that *'modified selection'* alters the shape of the fitness landscape (as shown in Figure 5.2).

As evolution smooths the fitness landscape to eventually become a single basin of attraction, the plasticity that moves the phenotype to the nearest optimum will more and more reliably discover the global optimum. Eventually, as the width of the single basin

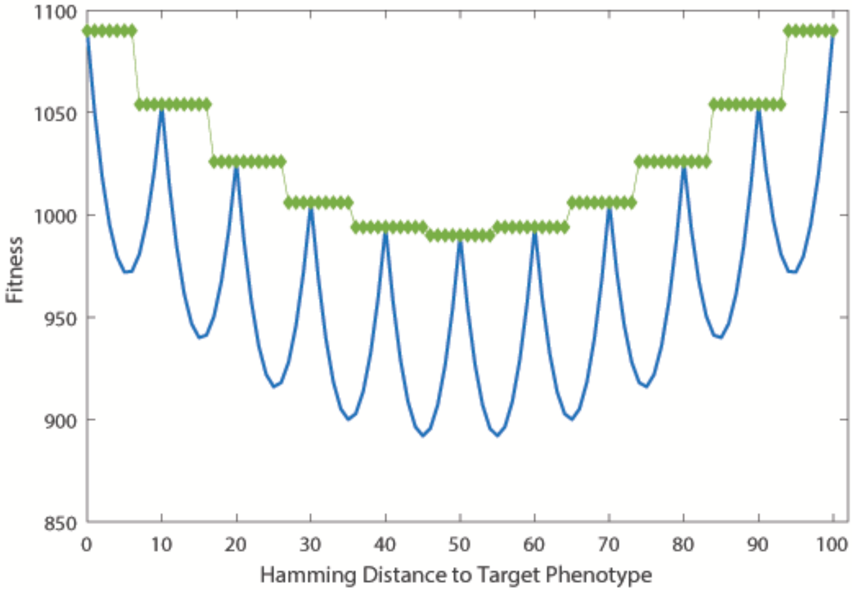


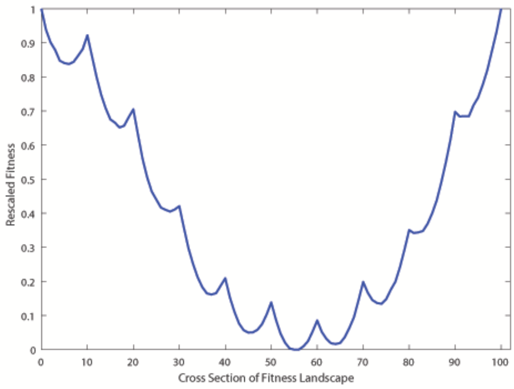
FIGURE 5.12: Cross-section of the fitness landscape with and without developmental plasticity. With developmental plasticity (shown in green with diamonds) and without developmental plasticity (shown in blue using a smooth line). The diagram shows a step by step transition between the target phenotype (Hamming distance = 0) and its complement (Hamming distance = 100). The action of learning moves the fitness to the nearest local (or global) optimum. See Section 4.1.1 for detail of how the cross-section of the fitness landscape is generated. Note: A p-value of 0.01 is used to aid clarity.

shrinks, a global optimal phenotype will be expressed regardless of environmental input.

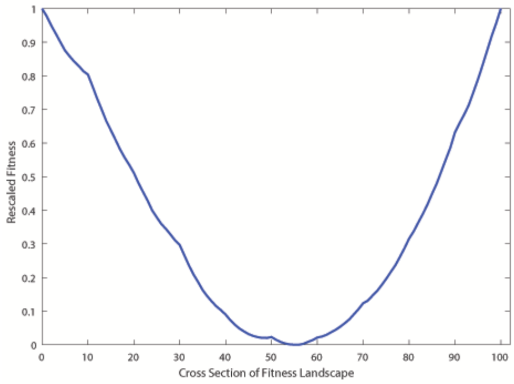
5.5.3 How Do Optimum Phenotypes Evolve Without Plasticity?

Although less reliably than with learning, as seen in both Figures 5.5 and 5.10ii, it is possible that, with a low enough mutation rate, evolution without learning can attain a globally optimum fitness more often than it would do by chance. This result is perplexing because, in the no learning case, the genotype is evolving to produce globally optimal phenotypes from random environments.

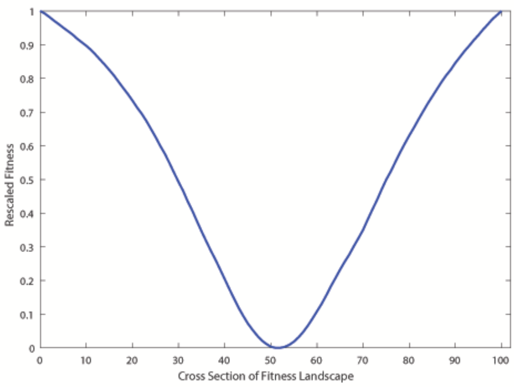
This issue potentially has relevance beyond the model presented in this chapter. The model structure: vectors of traits or environmental inputs interacting with matrices of correlations are a common way of representing and generating phenotypes in evolutionary simulations. Correlations between genes are often used as a representation of Gene Regulation Networks (examples include [121], [105], [30], [122], [33], [126], [59], [58]). For models where the vector represents part of the genome, these are usually evolved as an input to a development process, whereas where they represent environmental input, these are usually taken as start conditions to a development process [126].



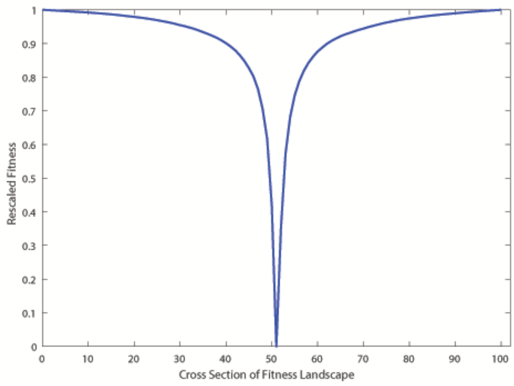
(I) Fitness landscape after 500 generations



(II) Fitness landscape after 2000 generations



(III) Fitness landscape after 4000 generations



(IV) Fitness landscape after 6000 generations

FIGURE 5.13: Evolution smooths the effective fitness landscape. The fitness landscape is generated at positions in the fitness landscape that transition step by step from one global optimum (at position 0) to the complementary global optimum (at position 100). The fitness landscape becomes progressively smoother as the correlation matrix evolves. See Section 4.1.1 for detail of how the cross-section of the fitness landscape is generated. Note in this assay, unlike the results shown in Figure 5.12 the phenotype is not thresholded before the fitness is measured but is rescaled so the shape of the landscapes can more easily be compared.

To explain how optimal phenotypes evolve from random inputs, we will consider a correlation matrix B encoding the correlation between n loci in a vector $\mathbf{s}$ where the entry B_{ij} specifies the correlation between traits s_i and s_j. To simplify the explanation, it is assumed the optimal fitness is achieved when all elements of the phenotype have a value of one, i.e. $s_x = 1, \forall s_{x \leq n}$ and only a single block of strong constraints in the MC problem is considered.

At first glance, it would appear illogical that there can be a signal from random vectors: when considering random input vectors that specify s_i and s_j, combinations of s_i and s_j occur equally often, mutations will be equally positive as negative and the number of ones (as opposed to negative ones) in the rest of the input vector are above and below 50% all ones equally as often. Therefore the direction of selection on s_i is equally often up as down. Consequently, one would expect the environmental inputs to cancel out and there be no consistent fitness signal to evolution.

However, there is a signal in the way random input vectors (e.g. the environmental inputs) interact with evolution due to the way loci within an input vector interact with the distribution of agreeing loci within an output vector (the phenotype).

For the case where the target phenotype is all positive[9], a 'correct signal' is defined as a change to the correlation matrix entry B_{ij} that has a positive fitness difference; this occurs when s_i and s_j are positively correlated and the sign of s_i and s_j is the same sign as the majority of all the other traits. Where s_i and s_j are negatively correlated, then a correct signal is when the sign of s_j agrees with the majority of all traits.

Whilst there is no correlation among input traits when inputs are random, the mutation to the correlation matrix isn't responding to correlations among two inputs - it controls the correlation between an input and the desired output (phenotype) - more specifically, the fitness-based selection on an output. And the desired output is not independent of the input because the input is part of the context of the output (and hence influences the direction of selection on the output). Consequently, there is a signal between inputs and outputs even when inputs are random.

To help illustrate this, consider the four combinations of the two loci (s_i and s_j) that can take either a positive or negative value with its fitness affected by a change in B_{ij} and also the context (the other values in the vector). Where the sign of s_i and s_j are of opposite sign (i.e. +ve and -ve or -ve and +ve), one combination gives the correct signal, the other will give an incorrect signal and therefore these two cases cancel out for any given context as they are equally likely to occur for any given Hamming distance to the target. Where a positive correlation agrees with the majority then there is a correct fitness signal but where it disagrees with the majority there will be an incorrect fitness signal. However, a positive correlation (i.e. +ve and +ve or -ve and -ve) that agrees with the majority in the input vector occurs more often than a positive correlation that agrees with the minority of ones (if we include s_i and s_j in the count of the majority or the minority). It is this differential frequency of positive correlations agreeing with the majority that provides the positive fitness signal in random environmental inputs.

Appendix B provides a calculation of the signal strength for the asymmetry described above which is derived to be:

$$Q = \sum_{z=0}^{n} X_z Y_z \tag{5.9}$$

[9]The same logic applies to any target configuration of loci within a block.

where X_z is the expected proportion of strings that have z correct loci - where a correct locus value is one that matches a target phenotype, in this case, a one and Y_z is the expected probability of correct signal for strings with z correct loci and

$$X_z = \frac{n!}{2^n z!(n-z)!}$$

and

$$Y_z = \frac{1}{n}\,|n-2z|$$

To check this calculation, the set of all random vectors of size n was enumerated along with the set of all possible positive and negative (non-cumulative) mutations on B. The number of 'correct' mutations was then plotted - where a correct mutation is a mutation in the direction of the sign of the correlation in the C matrix (in this example C contains all positive constraints). For this enumeration, in common with the ESDP function, the phenotype was subject to iterative development (as per equation 5.5) and the quadratic fitness function (as per equation 5.6). As can be seen by the graph in Figure 5.14 and the simulation results generated from random strings, the signal strength decays with n but the effect is still strong with a relatively large n.

As can also be seen from Figure 5.14, the iterative development function makes no difference to the enumerated result in this instance as the B matrix is initially blank and consequently a single mutation will have a very small effect on the trait values. Continuous values introduce additional stochasticity into the random vector input as compared to two-state vectors and therefore reduce the signal strength. The analysis in this section suggests that for any correlation model, where there is a random input, there will be a degree of skewing so that correct correlations will be sampled more often than incorrect correlations, even where there are continuous values. This explains why, in the non-plasticity / no learning cases of the correlation-based models presented in this thesis, there will be evolution of the on-blocks leading to the discovery of locally optimal phenotypes despite the input being essentially random. As correct correlations evolve, the signal strength from the random vectors will be substituted by the positive signal due to the correlations, so the effect is likely to be transitory. This analysis does not take into account the effect of weaker between-block constraints that are a feature of the MC problem, but the influence of the initial skew of random vectors is likely to be overtaken by the evolved within-block correlations before the between block correlations have an effect.

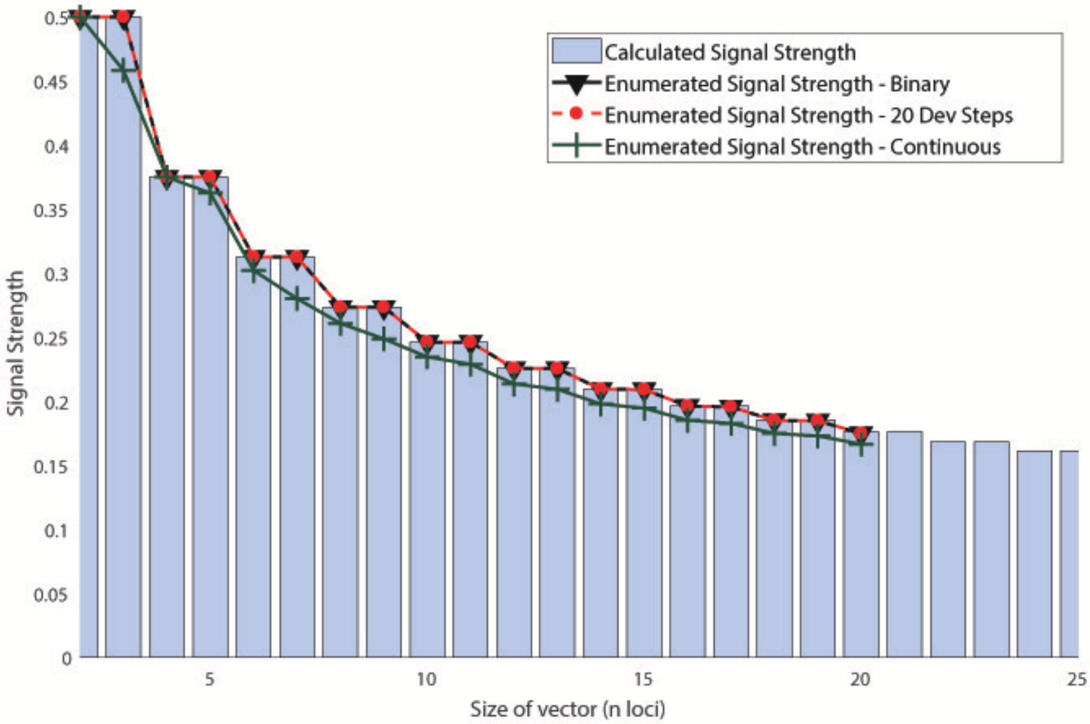


FIGURE 5.14: Signal strength by the size of vector for random vectors: The bars show the signal strength as calculated using equation 5.9 and this is verified as shown by the plot line by enumerating all possible random vectors of length n and testing with a positive and negative change to the correlation matrix (B) to discover the number of correct changes to B versus the number of incorrect changes to B. The black line with triangle markers shows random vectors of discrete values *1* and *-1*, the red dashed line with circle markers shows discrete values subject to twenty development steps and the dark green line with crosses shows continuous values with twenty development steps.

5.5.4 ESDP as a Model of Phenotypic Plasticity

There are two characteristics of the ESDP model that potentially may not be ideal as a model of developmental plasticity. The first is the assumption that phenotypic plasticity is always adaptive. The ESDP model is not intended as a model of the evolution of adaptive plasticity - the adaptiveness of the applied plasticity is a starting assumption and this is common to many models of learning and evolution (for example models by Borenstein [12], Frank [36] and Mayley [74]) but, admittedly is not so common in models of morphological plasticity. The assumed adaptiveness of the plasticity does have a logical basis: the model is based on the population mean, and where an individual's plasticity is deleterious, maladaptive plasticity would be strongly selected against and consequently would not be maintained in the population for long.

Secondly, adaptive plasticity is encoded as the temporary addition of fitness improving correlations during development using the X matrix. This is a relatively rudimentary, yet efficient method of making the phenotype's configuration at the end of development a function of the environmental input - which specifies the start condition to development - and an adaptive component. The outcome of this chosen method of encoding plasticity yields results similar to prior models of phenotypic plasticity. Figure 5.9 shows that when the B matrix is unevolved, this addition of fitness improving correlations (encoded in the X matrix) is moving the phenotype towards the local optima in a similar way that other models of plasticity move the phenotype's configuration to the local optima in 'trait space'; Figure 5.12 shows that with adaptive plasticity, development transforms the fitness landscape into a series of plateaus in a similar way to that of the Borenstein et al. model (as shown in Figure 3.1). It is worth noting that, at the end of evolution, the correlations matrix B is different from the constraints matrix C. For example, with the Concentric Squares problem, weights are evolved in the B matrix that are absent in the constraints matrix (as can be seen by comparing Figure 4.3 and Figure 5.8). However, it is acknowledged that the between-block values in the X matrix do cause some minor skewing towards the global optimum: where there is plasticity without evolution, the global optimum is encountered with a frequency of 0.82%, whereas a restart hill-climber only encounters the global optimum with a frequency of 0.23%.

5.6 Conclusion

This ESDP model demonstrates a model of the Baldwin Effect that relies upon canalisation of the genotype through epistatic effects to drive genetic assimilation. It is therefore significantly different from prior norm-of-reaction based models reviewed in Chapter 2 as the genetic and adaptive plasticity are operating in correlation space rather than the individual trait space. The ESDP model also shows that in a correlation model of plasticity, the discovery of high-fitness phenotypes through plasticity preceding a genetic canalisation through strengthening correlations accelerates evolution and, with noise in the fitness function or small fitness differences between local optima, allows more reliable discovery of phenotypes with optimal configurations.

This model also demonstrates an Expediting Effect, where evolutionary progress is accelerated by plasticity guiding evolution and, in turn, evolution smoothing the fitness landscape allowing plasticity to more regularly discover high fitness phenotypes.

In addition, we have shown that in this model, a cost of plasticity is not necessary and, unlike other models (e.g. Rago et al [97]), assimilation can take place even though the environment varies at each generation. Examination of conditions of assimilation is considered in more detail in Chapters 6 and 7.

Chapter 6

Correlated Behaviours Model of Learning

6.1 Introduction

As discussed in previous chapters, most models of learning and evolution represent learning as either hill-climbing on the phenotype to a local or global optimum (e.g. Hinton & Nowlan model as discussed in Section 2.3.4) or as a range of phenotypic values dictated by the environmental input (e.g. Papaj [90]). In this way, learning produces higher fitness phenotypes, with potential for the genotypes - via a selection pressure on existing genetic variability or by mutation - to 'catch-up' with the learnt phenotype. Even though these models of learning and evolution use evolution to bound the space that in-life learning searches - for example the Hinton & Nowlan model substitutes flexible alleles for fixed alleles - they do not use what is evolved to guide what is learnt: the search of the fitness landscape by learning is independent of the innate behavioural traits. Furthermore, both the hill-climbing and span of plasticity representations used in previous models treat behavioural traits as individual, uncorrelated entities and, as discussed previously in Section 4.2.1, the work of Lorenz [69] and others suggests that learning utilises combinations of low-level innate behaviours.

In addition, many prior models of learning and evolution appear to be based on the premise that innate behaviours are static - often, the start point of the learning is the genotype which is modelled as insensitive to environmental input (e.g. Borenstein et al. [12] and Mayley [74]). As discussed extensively in Chapter 2, innate behaviours are subject to environmental stimuli: the goose needs the stimulus of the egg (or maladaptively the beer bottle) being outside the nest to trigger it to roll the object back into the nest. What differentiates innate behaviours from learnt behaviours is the ability for previous experience to affect the behavioural outcome and the ability for that outcome to be retained or lost between generations. This aligns with Mery & Burns [77] view of innate behavioural plasticity as a combination of learnt and innate behavioural responses where both types of behavioural expression are environmentally sensitive. And, unlike many models of learning and innate behaviours that use binary representations to encode behaviours, the representation of the behavioural phenotypes presented here can change the level of expression on a continuous scale. Consequently, a model that allows learning to use coordinated sets of innate behaviours and in which both learning and evolution are environmentally sensitive is likely to have an interesting and novel dynamic that is distinct from previous models of learning and evolution.

One potential insight into a dynamic between coordinated traits and learning was West-Eberhard's [129] hypothesis that the mutual feedback between learning and the evolution of physical traits is likely to facilitate the rediscovery of learnt behaviours (as discussed in Section 2.7.4). West-Eberhard uses the example of the evolution of big beaks and certain gut morphologies rewarding the learnt behaviour of choosing to eat larger, more nutritious nuts. This, in turn, reinforces selection for a larger beak size and gut morphology, thereby coordinating their evolution and creating a positive feedback loop. We now build on this concept by considering how learning and innate behaviours can form a similar feedback mechanism where learning coordinates innate behaviours which then reinforces the selection of those innate behaviours together. These innate behaviours then constrain learning, reducing the phenotypic search space so that the learning is focused on high fitness behavioural combinations.

Using the bee foraging conceptualisation described in Section 4.2.1 to illustrate this two-way feedback: based on environmental input, a bee will exhibit a set of foraging actions (scouting, resource selection, resource collection, etc.) which are a combination of innate and learnt behaviours. Learning will test combinations of innate behaviours, for example choosing to collect pollen or nectar, selecting red or blue flowers and performing sonication and blossom tripping. Given an environmental input, the best combination of behaviours discovered by learning will determine the fitness of the bee[1]. As learning finds good combinations of individual innate behaviours, selection can reinforce the linkage between those behaviours - for example, genetically correlating the behaviours of pollen collection, choosing red flowers and sonication. As these behaviours become increasingly genetically linked, the learning is more and more likely to use those good combinations of behaviours together and in doing so makes the discovery of ideal combinations for a given environmental input more probable.

We include this bi-direction feedback between learning and innate behaviours into this Correlated Behaviours model; learning directly alters the expression of innate behaviours and genetic connections between innate behaviours constrain what can be

[1]It is recognised that the fitness benefits of the bees behaviour may accrue to the colony

learnt. Consequently, this simulation follows the causal process flow as set out in Figure 2.4 (B), but expanded so that the model tracks the interactions between innate behaviours and learnt behaviours; evolution's impact on non-behavioural traits is not considered. Thus, the flow being modelled is as shown in Figure 6.1.

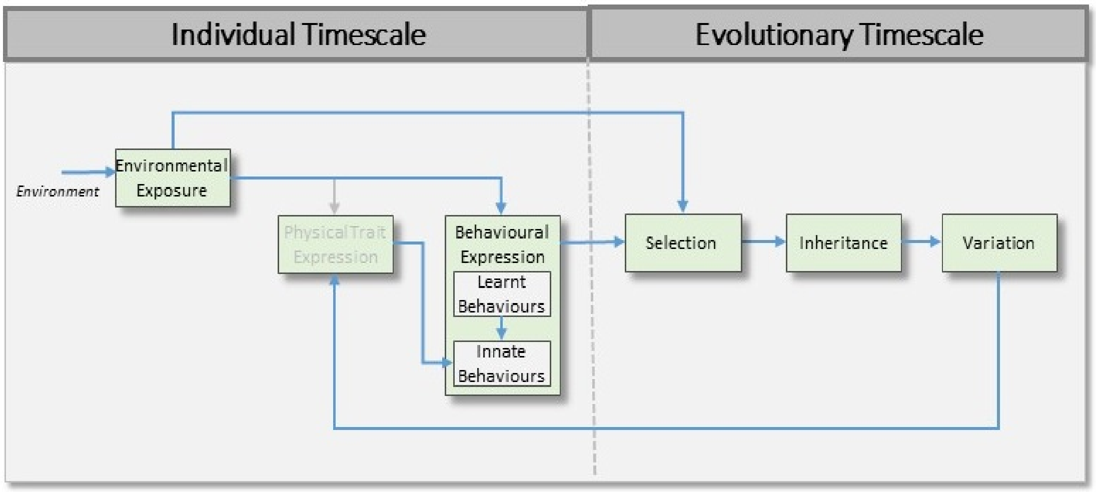


FIGURE 6.1: In the Correlated Behaviours Model, behavioural expression is triggered by environmental stimuli and is a combination of innate and learnt behaviours. This combined behavioural expression has fitness consequences that directly alter the selection on the phenotype. Variation changes the strength of connections between innate behaviours which alters the probability that they are used together by learning. Physical trait expression (box with greyed text) is assumed to be static throughout an individual's lifetime and evolution and is therefore not explicitly modelled.

In common with the ESDP model, the increasing linkage between genetic traits - in this case innate behaviours - canalises learning, so behaviour becomes increasingly environmentally insensitive. This follows a model of assimilation that both Crispo [22] and Loison [68] argue is closer to Waddington's [119] original intention for the action of genetic assimilation: increasingly strong genetic connections between traits make the resultant phenotype increasingly resilient to environmental perturbation. This concept can also be applied to learning as Tierney [116] citing Steddon suggests that innate behaviours can be viewed as a form of canalisation of learnt behaviours, with varying degrees of flexibility to be modified by learning. They go on to argue that learning is a form of plasticity that gets genetically assimilated (these ideas are discussed more in Section 4.2).

As a model of learning - unlike the ESDP model that deterministically used correlations to express plasticity - there is a trial-and-error element associated with operant learning. Here, we allow learning to trial combinations of innate behaviours, where the stronger the genetic link between the innate behaviours, the more likely those behaviours are to be expressed together in a direction dictated by those genetic correlations. A random threshold determines the connection strength required for innate behaviours to be trialled together and because this threshold changes randomly at each trial, there is a stochastic element to the learning.

The mechanism for achieving this co-expression of innate behaviours mediated by learning is inspired by Watson, Mills and Buckley's [123] generative dynamical system model (rHN-G[2]), the detail of which is described in the next section. This generative system allows for a search to scale up, first finding good combinations of individual behavioural traits and then finding combinations of sets of traits. An adapted form of this generative model of learning works well as a model of reinforcement learning because it enables genetic correlations to constrain learning and allows learning to act on increasingly sophisticated combinations of innate behaviour.

Rather than smoothing the fitness landscape as is the case with the ESPD model, in the Correlated Behaviour model, the coordinated changes to the learnt phenotype's innate behavioural expression allows movement across fitness valleys, as shown in Figure 6.2

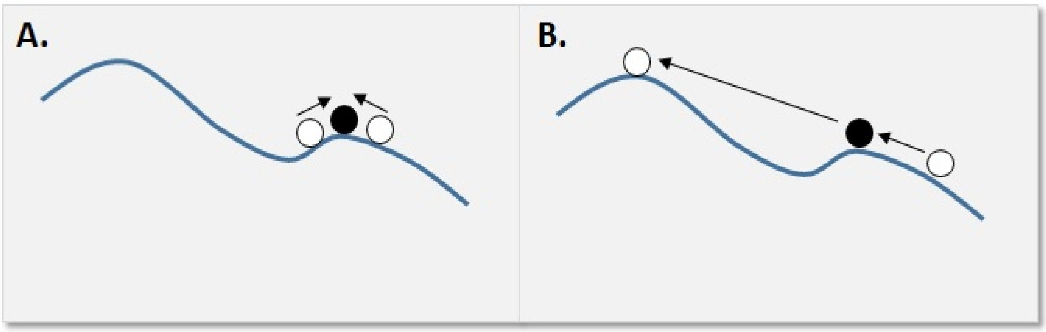


FIGURE 6.2: Adapted from Watson, Mills & Buckley [123]. (A) A phenotype (black dot) is stuck on an adaptive peak because it has a higher fitness (denoted by black arrows) than its single mutation variants (white dots). (B) By coordinating change to multiple innate behaviours based on genetic linkage, the phenotype can escape the local optimum (denoted by black arrows).

This, therefore, models a different form of canalisation: genetic linkages are still constraining the production of phenotypes but the constraints change the areas of phenotype space that can be searched. As an example, when all traits within blocks correctly correlate with all others in their blocks, the search of phenotype space is constrained to local optima.

As will be seen, this novel experimental set-up demonstrates that with learning, a behavioural phenotype with globally optimal fitness can quickly be discovered, whereas, without the benefit of learning, the behavioural phenotype takes a long time to find locally optimal phenotypes.

A key attribute of this model is that it does not contain a cost of learning or a learning outcome that is consistent across generations. Consequently, the conditions under which assimilation of the learning can take place are significantly relaxed.

[2]The initials stood for restart Hopfield Network - Generative associations

6.2 Model Structure

As previously stated, the structure of this model is broadly based on Watson, Mills & Buckley's [123] rHN-G concept. The original rHN-G model used a random threshold and a correlation matrix to determine how to change a solution vector where correlation strengths are updated deterministically based on Hebbs rule. However, to be a model of learning and evolution, this concept has been expanded in three important respects.

Firstly, to ensure that co-expression of innate behaviours is subject to **trial-and-error learning**; different combinations of innate behaviours are iteratively tested together dependent on whether their correlation strength is above a random threshold. This threshold changes at each learning trial, with the combination of behaviours being retained if that change leads to a behavioural expression that is fitter than the previous trial. Lifetime fitness for the purposes of selection is the fitness at the end of a defined number of learning trials. As a consequence of the mutation of genetic correlation strengths between innate behaviours, the innate behaviours are more or less likely to be tested together by learning. So that the coordination of behaviours orchestrated by learning has a strong effect, the direction of the expression of the behaviour during learning is determined by whether the behaviours are negatively or positively correlated in the correlation matrix (i.e. the sign of the correlation). This simplification essentially makes learning a directional choice of behaviour, e.g. red or blue flowers, pollen or nectar, blossom tripping or sonication (we call this *'trait flipping'*). For comparison, we model the no learning case as each learning trial happens but there is no test to see if each that trial has produced a fitter behavioural phenotype. Consequently, if there are strong and consistent correlations between innate behaviours, behavioural expression still has the potential to change multiple traits in unison but, in this case, will do so at random.

Secondly, mutation and selection of the resultant behavioural phenotype are used to drive an **evolutionary process**: under the SSWM assumption, a population mean, unmutated behavioural phenotype is compared to the mutated behavioural phenotype with the fitter retained. This comparison can be made for both the with and without learning scenarios. In common with the ESDP model, the main motivation for using the SSWM assumption is to make the simulating both in-life learning and thousands of generations of evolution computable within a reasonable timescale. However, this assumption adds complexity to the model as it requires the population mean and the mutated population mean to have equivalent learning. Consequently, the random elements that allow learning to determine which innate behaviours are expressed together during a learning trial need to be consistent. These random components are computed at the beginning of every generation and applied by learning. This is a less strict assumption than Mayley's [74] model wherein the selection of behavioural trait values that could be altered by learning is determined at the start of evolution and remains static across all generations[3].

Thirdly, for evolution to reward fitter behaviours, the **behavioural expression** with and without learning needs to provide a fitness signal that distinguishes between the mutated and unmutated population mean behavioural phenotype. There is potential for learning to mask this fitness signal (as discussed by Mayley [73] and in Section 2.4). So, in this model, the strength of correlation is also used alongside the environmental input to determine the strength of the expression of a particular set of innate behaviours (as well as the direction): the strength of expression of each behaviour is a function of the magnitude of the environmental input for a particular trait and the correlation strength between innate behaviours. For a particular trait, if connections to other traits are low, then the effect of its own trait values - that are initially the magnitude of the environmental input - on those other traits will also be low. This expression is dependent on all correlated traits, not only those above the random threshold and the strength of expression of each behaviour is recalculated for each behaviour trial, so between learning trials, learning remembers fit combinations of behaviours rather than the strength of the expression of those behaviours.

Simply allowing innate behaviours to have genetic linkage explicitly provides learning with the capacity to scale up: as innate behaviour connection strengths increase, the chances of behaviours being tested together by learning increase. And so, learning moves from trying combinations of single innate behaviours (no evolved connections) to sets of behaviours being tested together with a high probability (strong evolved connections). Consequently, connections between innate behaviours progressively constrain the degrees of freedom of what can be learnt. Ultimately, if all innate behaviours are fully connected either positively or negatively, learning will be redundant because the optimum behavioural configuration will be fully specified by the innate behaviours: learning will not increase phenotypic fitness because the phenotype without learning will express a phenotype with optimum fitness.

In summary, with this structure, correlations between innate behaviours are used in three ways; (1) they determine the probability as to whether innate behaviours are tested together, (2) where learning alters behaviours, the sign of the correlation is used to determine the direction of a behaviour and (3) the strength of connections are used to determine the strength of behavioural expression for a given environmental input.

Both the Modular Constraints (MC) problem with its building-block structure and the Concentric Squares (CS) problem with its localised constraints are used as test problems.

[3] Mayley's population model [74] did allow different individuals to exhibit different learning although the population was highly converged. See Section 3.3.2 for details.

6.3 Model Detail

In common with the ESDP model, the Correlated Behaviours model uses a matrix of genetic correlations (B) subject to mutation. In this model, the B matrix represents correlations between innate behaviours[4]. This matrix is used to determine which innate behavioural traits are likely to be expressed together by learning. The behavioural phenotype, which is a combination of innate behaviours changed by learning, is subject to environmental input and the fitness function which in turn is defined dependent on the test problem being assessed. The environmental vector ($\boldsymbol{\epsilon}$), correlation matrix (B), mutation matrix (J), fitness function (w) and fitness thresholding function (θ) are the same as that for the ESDP model, provided by equations 5.4, 5.1, 5.2, 5.6 and 5.7 respectively. The detailed relationship between model functions is shown in Figure 6.3 and is further defined in the following sections.

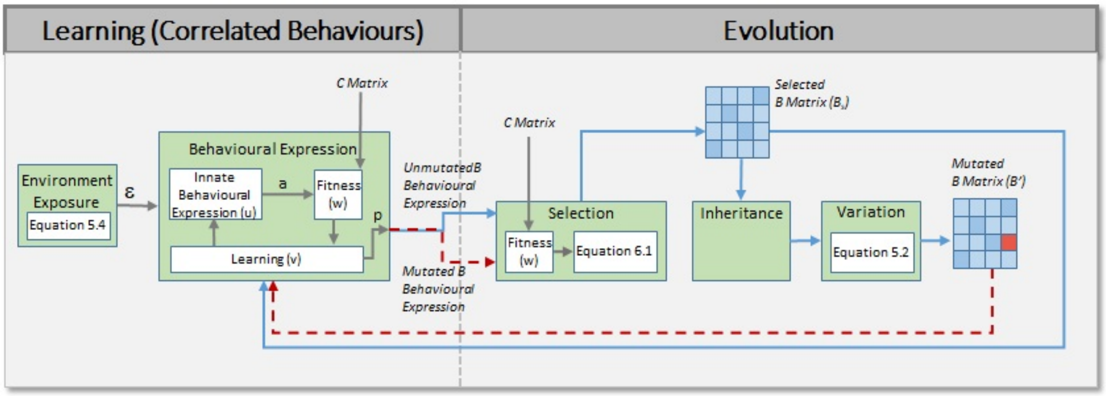


FIGURE 6.3: Following the structure shown in Figure 6.1, an illustration of the interaction between Correlated Behaviours model functions. The learning within an environment $\boldsymbol{\epsilon}$ is assessed for both the unmutated (B) and mutated (B') innate behaviours correlations matrix using the learning function (v), and an innate behavioural expression function (u) and the fitness function (w). The learning function (v) iteratively tests different combinations of innate behaviours by generating a trial behavioural phenotype $\mathbf{a}$ and testing if is fitter than the previous trial using the fitness function (w). At the end of learning, the behavioural phenotype ($\mathbf{p}$) is the behaviour used for the evaluation of lifetime fitness. On the evolutionary timescale, the mutation to the B matrix is generated using equation 5.2 (as per the ESDP model) and equation 6.1 is used to select whether, after learning and innate behavioural expression, the mutated (B') or unmutated (B) correlations matrix is inherited to become the next generation (again using the fitness function w and the C matrix). The flow of the unmutated form of the correlation matrix (B) and resultant developed phenotype is denoted by solid blue lines whereas the flow of the mutated correlation form of the correlation matrix (B') and derived developed phenotype is represented by red dashed lines. It should be noted that, in common with the ESDP model, the selective environment is defined by the C matrix which varies dependent on problem type.

[4]Rather than a driver of a physical trait development function, as is the case with the ESDP model.

6.3.1 Evolution

Under the SSWM assumption, evolutionary change is represented by mutations to the correlation matrix (B) and testing whether the mutation produces fitter behavioural phenotypes - both with and without learning.

At each evolutionary step, a mutation matrix, which represents a single positive or negative mutation to a random correlation is added to the B matrix to provide the mutated correlations matrix (B'). If the mutation increases the fitness of the phenotype, it is retained, otherwise, it is discarded. Unlike the ESDP model, in this model, the values of the B matrix are capped at values between -1 and 1. Therefore whether a mutation to B is accepted is given by:

$$B_{(s+1)} = \begin{cases} B_{(s)} + \gamma J^{kl}, & \text{if } w(v(B_{(s)} + \gamma J^{kl}, \boldsymbol{\epsilon})) \geq w(v(B_{(s)}, \boldsymbol{\epsilon})) \wedge |B_{(s)_{i,j}}| \leq 1 \\ B_{(s)}, & \text{otherwise} \end{cases} \tag{6.1}$$

In the above, γ defines the magnitude of the mutation, v is the learning function (defined below), w is the fitness function as defined previously by equation 5.6 and $\boldsymbol{\epsilon}$ denotes the environmental input vector. The sequence of mutations for that generation completes at $B_{(s=n)}$; the number of mutations of B per generation is defined by the number of genes in the genotype. The magnitude of the mutation is bounded by $\pm d$, and the mutation is uniformly distributed where $\gamma \sim U([+d, -d])$. In the experiments using this model, d is relatively large as compared to those used in other correlation models. This is because for a learning trial to be different from previous trials, there needs to be a good chance that the threshold is above some correlation values and below others. If mutations are small and of a similar magnitude then there is a low probability of this occurring. Unless otherwise stated, d is 3x10^{-3}.

To expose the population to a range of environmental conditions, the environmental input is changed at the start of each generation (but not during every test mutation of B).

6.3.2 Learning (v function)

The learning function v is defined in the following way. A behavioural phenotype $\mathbf{p}$ of n behaviours is subject to random environmental inputs $\boldsymbol{\epsilon}$, so that at the start of its lifetime a behavioural phenotype is equal to the environmental inputs:

$$\mathbf{p}^{t=0} = \boldsymbol{\epsilon}$$

where

$$\mathbf{p} = (p_1, p_2, \ldots, p_n) \in \mathbb{R}^n$$

The behavioural phenotype changes throughout its learning phase by trialling combinations of behaviours based on the strength of genetic correlations; a trial combination of behaviours is specified by the vector **a**.

$$\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$$

As was the case in the original rHN-G model [123], for a maximum number of behavioural trials η, at each trial t, a focal behaviour c is chosen at random for every learning trial.

$$c \sim U(\{1, 2, \ldots, n\})$$

For experiments in this chapter, unless otherwise stated η is set to 300 - this number of trials helps ensure that at the start of evolution there is a reasonable chance that learning will produce a phenotype that is close to a local optimum. The direction of expression of each trial set of behaviours **a** is based on the correlations in the B matrix. If the correlations between innate behaviours encoded in the B matrix are above a random threshold (r), the sign of the element of **a** is set according to the sign of the correlation matrix entry (in B) and the sign of the focal behaviour ($\mathbf{p_c}$). The magnitude of the trial behavioural expression starts as the same as the corresponding innate behaviour ($\mathbf{p}$). If the correlation entry in B is below the threshold, then trial behaviour is the same as the behavioural phenotype $\mathbf{p_y}$. This has the effect of creating a trial set of behaviours by changing the direction of expression of individual behaviours that are above the threshold (r). So the trial behaviour generated in a learning step is defined as:

$$a_y^t = \begin{matrix} -sgn(p_c^t B_{cy})|p_y^t|, & |B_{cy}| \geq r \\ p_y^t, & \text{otherwise} \end{matrix} \tag{6.2}$$

where the threshold is a random magnitude between the average of the absolute magnitudes of entries in B and the maximum magnitude and:

$$r \sim U(\alpha, 1), \alpha = \frac{\sum_{i,j}^{n} |B_{ij}|}{n} \tag{6.3}$$

The part of the model has additional complexities driven by the SSWM assumption and the representation of the population as a population mean phenotype and population mean genotype. So that there is consistency between the comparison of B and B', the set of random thresholds r, focal traits c used during the learning trials are predefined at the start of each generation.

With learning, the trial behaviours become the new behavioural phenotype if the fitness of the expression of the trial set of behaviours is greater than the fitness of the expressed behavioural phenotype prior to that trial (the fitness increasing behaviours are remembered between learning trials).

$$\mathbf{p}^{t+1} = \begin{cases} \mathbf{a}^t, & w(u(B,\mathbf{a}^t)) \geq w(u(B,\mathbf{p}^t)) \\ \mathbf{p}^t, & otherwise \end{cases} \tag{6.4}$$

where, u is the behavioural expression function and w is the fitness function defined by equation 5.6.

Without learning, the expression of linked innate behaviours happens at each trial but the combination is always retained, so there is no opportunity to discard unfit behaviours. In this way, the phenotype's fitness is always determined by innate behaviours: $u(B,\mathbf{a})$. To ensure that in the absence correlations, apart from self connections, $\mathbf{p}^{t=0} = \boldsymbol{\epsilon}$, the initial instantiation of the correlations matrix (B) is the identity matrix ($B_{(i,i)} = 1$ and $B_{(i \neq i)} = 0, \forall i$).

So that at the end of the learning trials there is a fitness difference between the mutated and unmutated B, the behavioural expression of the learnt phenotype is returned as

$$v(B,\boldsymbol{\epsilon}) = u(B,\mathbf{p}^{t=\eta}) \tag{6.5}$$

6.3.3 Innate Behavioural Expression (u function)

The behavioural expression function u defines the strength of behavioural expression of the phenotype and takes the simple form:

$$u(B,\mathbf{p}) = \mathbf{p}B \tag{6.6}$$

In this way, the strength of expression of the behavioural expression of each trait is a function of the trait's environmental input (as $\mathbf{p}^{t=0} = \boldsymbol{\epsilon}$) and the inputs of the traits that it is correlated with.

6.4 Results

6.4.1 Evolution with Learning Consistently Finds Globally Optimal Phenotypes

The first test of the Correlated Behaviours model is to compare the performance of evolution of the phenotype with and without learning for the MC problem. For an initial comparison, the same parameters for the learning and no-learning case are used to assess the reliability of the phenotype reaching the globally optimal fitness across 30 runs of the model.

As shown in Figure 6.4, evolution together with learning consistently discovers phenotypes with a globally optimal phenotypic fitness. It is worth noting that although the number of generations needed to find the globally optimal fitness is approximately the same as the ESDP model presented in the previous chapter (circa 10,000), the number of mutations per generation is approximately one-hundred-fold fewer and therefore this model is reaching a global optimum with significantly fewer evolutionary fitness evaluations that the ESDP model. Whilst the number of fitness evaluations by selection is less, the number of fitness evaluations by learning is significantly more; there are no fitness evaluations within a lifetime in the ESDP model.

Unlike the ESDP model, the no plasticity (i.e. no learning) case does not appear to have an appreciable increase in fitness. However, analysis of the average number of correct mutations shows that for the no-learning case, there is a small signal to evolution as there was in the no-plasticity case for the ESDP model. This is tested by altering the parameters so effects from a small fitness signal can be observed within a reasonable timescale: the no-learning case is run for an extended number of generations (250,000) as shown in Figure 6.5.

The extended run of the no-learning case shows that for the first 130,000 generations there is an appreciable increase in fitness. At approximately 130,000 generations, most fitnesses are locally optimal and the global optimal fitness is encountered on occasion - this continues until the end of evolution and remains irregular.

To confirm that the expressed behavioural phenotypes are not reliably producing the same local or global optimum, for one run of the same experiment shown in Figure 6.6, the red squares indicate where a phenotype with the same fitness is produced consecutively for two generations. This analysis suggests whilst the no learning case produces locally optimal phenotypes, there is canalisation between individual behaviours to produce locally optimal phenotypes but there is no effective evolution of the between-block correlations and so the global optimum behaviours are not reliably produced. Where the number of consecutive generations is raised to four, there are no red squares

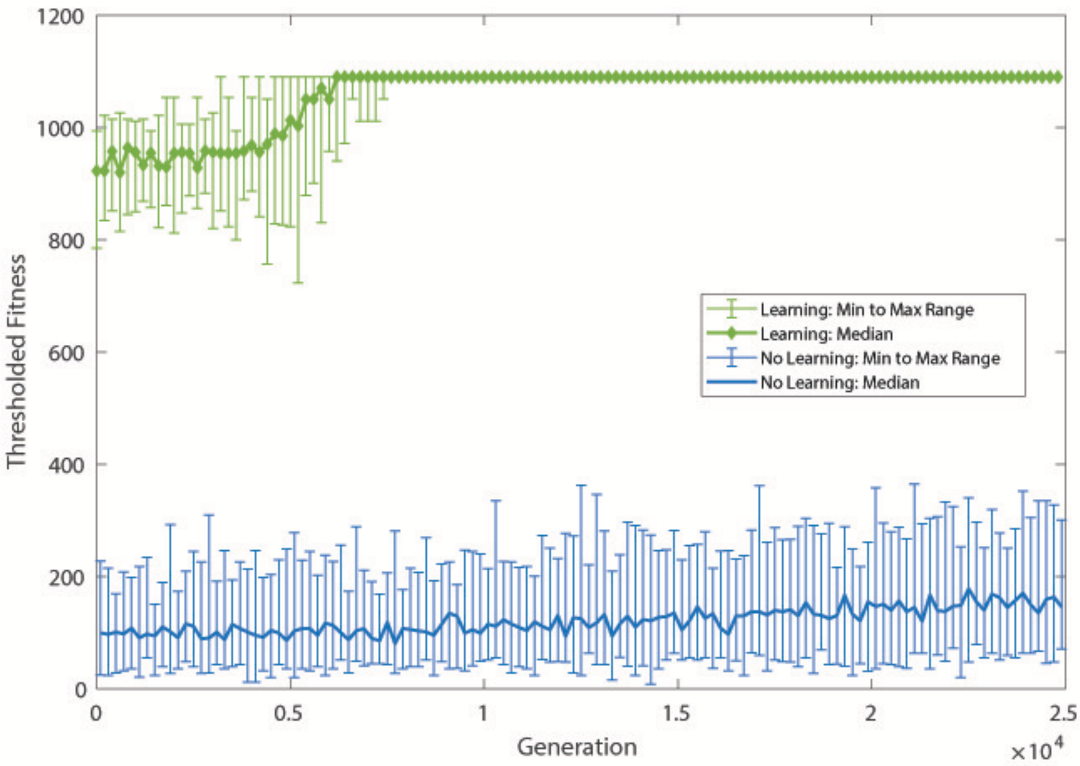


FIGURE 6.4: The median fitness of the population average phenotype without learning is tracked at every 200 generations for 30 runs of the MC problem. The green line with diamonds shows the fitness of phenotype with correlated learning, whilst the blue line shows evolution without learning. The bars show the minimum and maximum fitness achieved.

produced (not shown) and so, without learning, the same phenotypic fitness is never produced four generations in a row.

To understand why this is the case, it should be remembered that even without learning there is behavioural expression of correlated innate behaviours, but without a test to see if fitter expressions should be retained. Therefore, at the start of evolution, the expressed behaviours are random, but as explained in Section 5.5.3 there is still a small fitness signal towards correct correlations within a block. Therefore, over many generations the within block correlations form and the expressed behaviours become locally optimal. Once those within-block correlations form, no-learning behavioural trials flip blocks of innate behavioural traits, essentially making the behavioural expression random at the block level. Consequently, the relationship between sets of highly-correlated behaviours is random with respect to the environmental input.

Therefore, a Baldwin Optimizing effect can be observed: with learning, globally optimal phenotypes are discovered and canalised whereas without learning only locally optimal phenotypes are produced due to the canalisation of local connections.

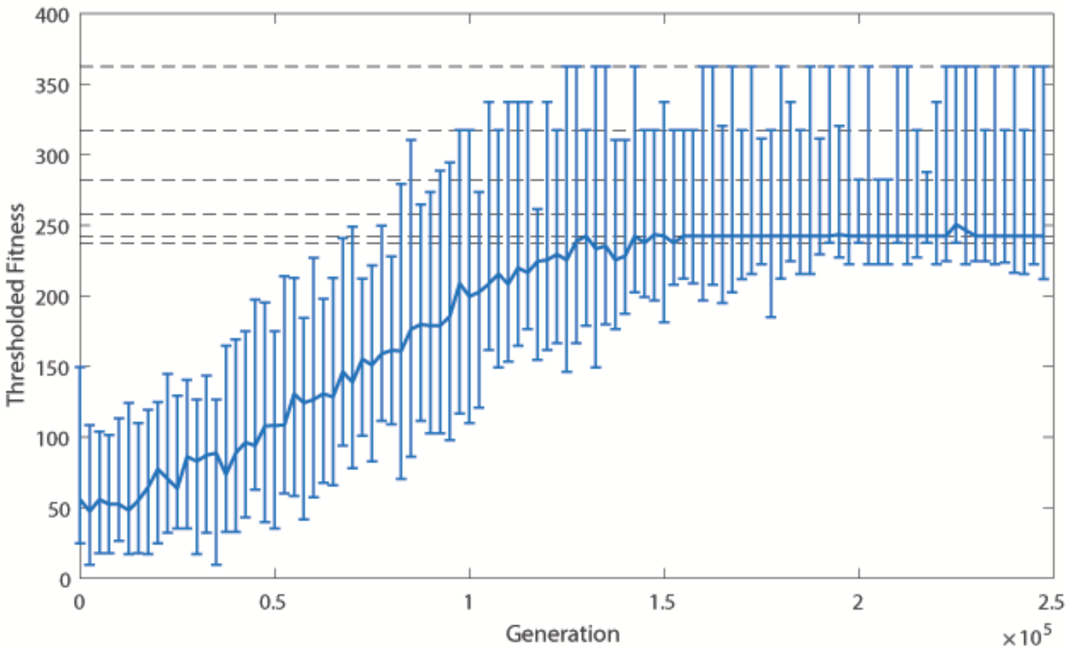


FIGURE 6.5: The median fitness of the population average phenotype is tracked at every 2500 generations for 30 runs of the MC problem without learning. The blue line shows the median fitness of phenotype and the bars show the minimum and maximum fitness achieved. The black dotted line represents the locally optimal fitnesses for the landscape which is generated using ten blocks of size five. The parameters have been tuned so that the effect can be observed in a tractable timescale (10 blocks of size 5, p=0.05, $d = 2 \times 10^{-3}$, $\eta = 200$).

6.4.2 Learning Scaling from Coordinating Single to Multiple Traits Further Accelerates Evolution

To validate the mechanism by which the with-learning results in Figure 6.4 occur, we analyse the action of learning by assessing how many traits are changed during learning. Figure 6.7 (I) shows that at the start of evolution, learning is changing approximately one innate behavioural trait during a learning step, suggesting that learning is hill-climbing to a local optimum. As evolution progresses, behavioural traits become increasing genetically correlated and therefore learning tests a range of behavioural traits together (dependent on the random threshold being above or below the correlation strength for a given learning step). By the end of evolution, during a generation, a maximum of approximately nine traits are flipped during a learning step. Panels (II) to (IV) of Figure 6.7 show a diagrammatic representation of the values in the correlations matrix (B) at different stages of evolution for one run of the simulation. The pattern of correlation values suggests that at the start of evolution, learning is changing the direction of individual behaviours, but as evolution progresses learning begins to change the direction of multiple, closely correlated behaviours. We, therefore, conclude that learning has scaled-up from flipping a single trait to changing the sign of multiple, strongly correlated traits.

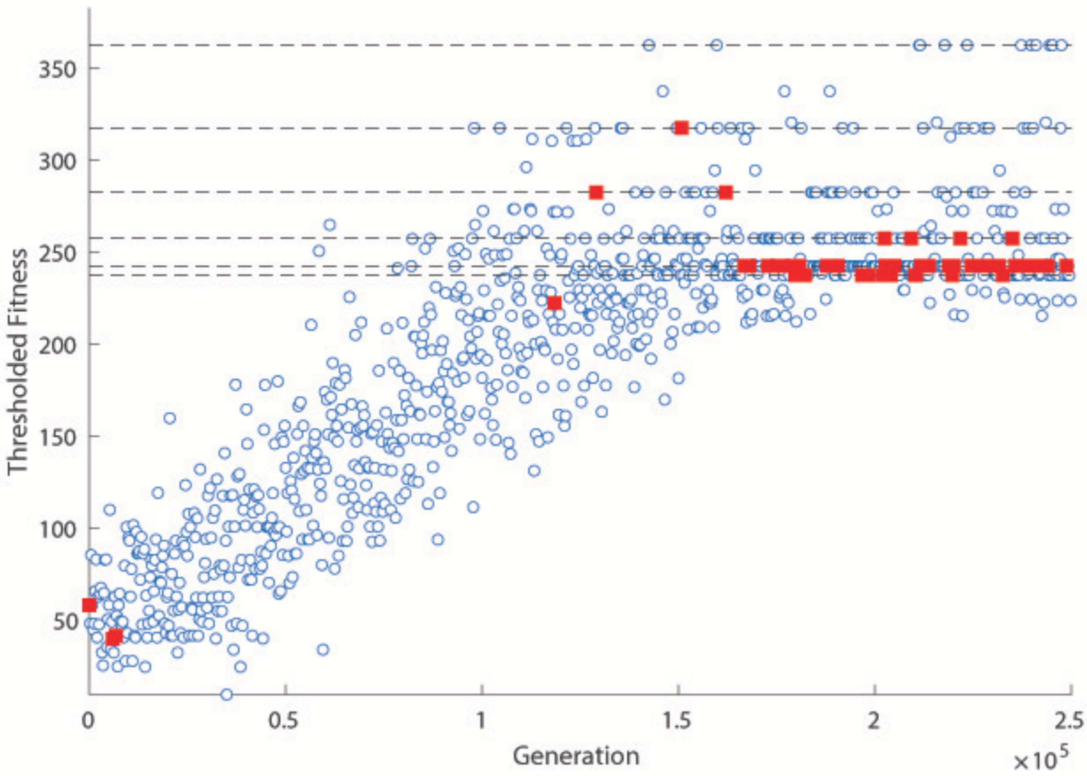
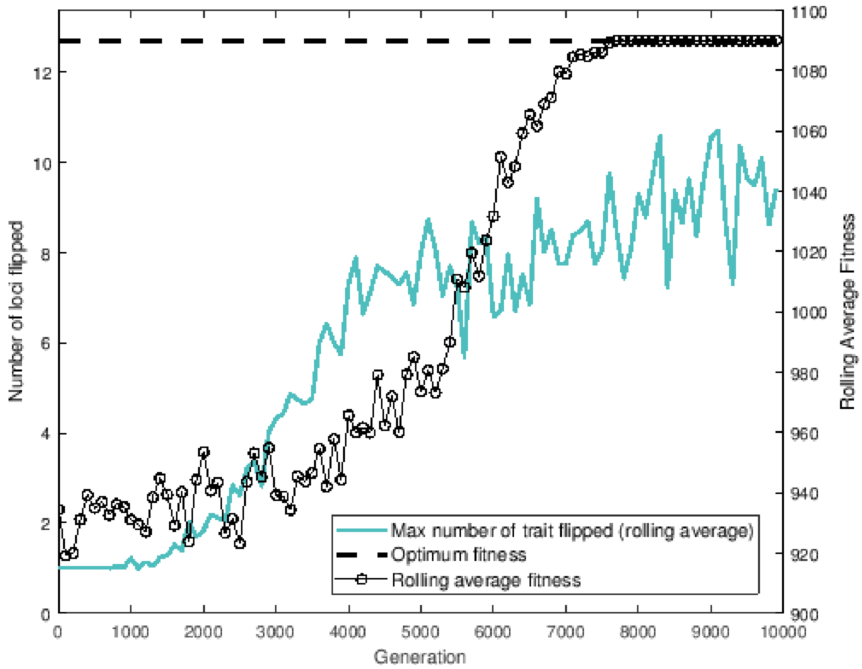


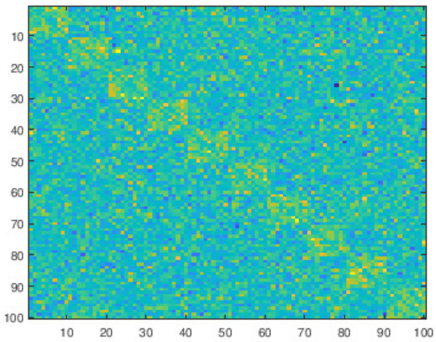
FIGURE 6.6: For one run of the same experiment as per Figure 6.5 the fitness is sampled every 250 generations (shown by the blue outlined circles). Where the sampled and the immediate next generation produces the same phenotypic fitness is shown by red squares.

This scaling-up also appears to accelerate the rate of evolution, as also indicated in Figure 6.7 (I). As the maximum number of traits flipped per generation increases, the rate of fitness change also appears to increase: between the first generation and generation 2,000, single trait changes occur and the fitness of the phenotype climbs slowly. Between generations 2,000 and 5,000, the number of traits changed in unison increases to approximately eight (i.e. most of a block of 10), and over the same period, the rate of change of phenotypic fitness rises markedly. Phenotypic fitness continues to improve at this faster rate until the globally optimal fitness is achieved.

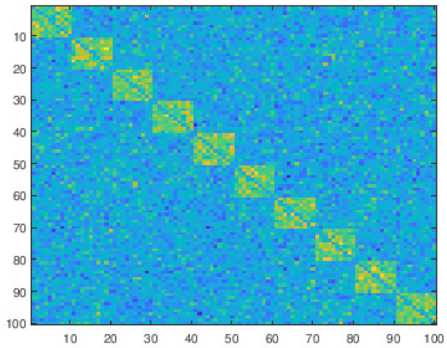
These results in conjunction with the correlation matrices shown in Figure 6.7 suggest that for fitness improving learning steps most of the traits within one single block are being flipped within a single learning operation, rather than multiple blocks changed at the same time. Consequently, the between-block correlations which are seen to evolve in (V) of Figure 6.4 are probably not the key driver of evolutionary progress towards a global optimum in this model. This is tested in the next assay.



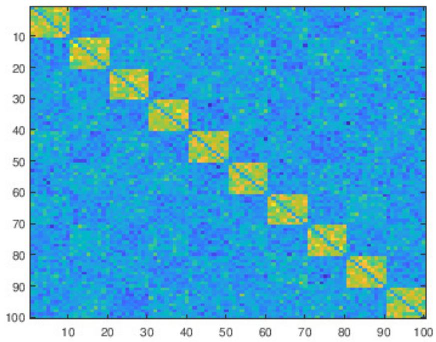
(I) Rolling average maximum number of behavioural traits flipped by a fitness improving learning step.



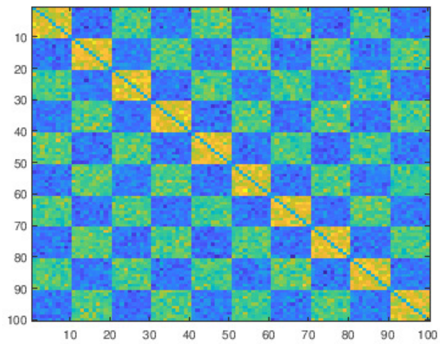
(II) 1000 Generations



(III) 2500 Generations



(IV) 5000 Generations



(V) 10000 Generations

FIGURE 6.7: (I) The maximum number of behavioural traits flipped by a fitness improving learning step for the last mutation of each generation is averaged over 50 generations (blue line). As the number of traits flipped increases, the fitness of phenotype rapidly increases (fitness is shown by the black line with circle markers). By the end of evolution, learning flips approximately 9 to 10 traits and maximal fitness is achieved. (II) to (V) The B matrix at selected generations. Values are shown as a continuum of colours, blue (negative) to yellow (positive) values. Note the values of $B_{i,i}$ have been removed so that the differences between evolved correlations show more clearly. Strongly correlated blocks are evolved before weakly correlated blocks.

6.4.3 Learning without Evolution of Between-block Correlations

To help assess to what extent the evolved between block correlations are a driver to reaching the global optimum, we compare the evolution of both the within and between block correlations with an assay where between block connections cannot be evolved using the MC problem. All other operating parameters remain constant. Figure 6.8 shows that when evolution is restricted to within block correlations only, learning can find globally optimal phenotypes (shown by the brown line with small dots) but the progress of evolution is slower than if the between block connections are allowed to evolve (shown by the green plot with diamond markers). This suggests that learning can flip multiple blocks using the fitness differences driven by the p-value in the fitness function but evolution with learning does not need to genetically correlate these weakly linked behaviours to reach optimum phenotypes.

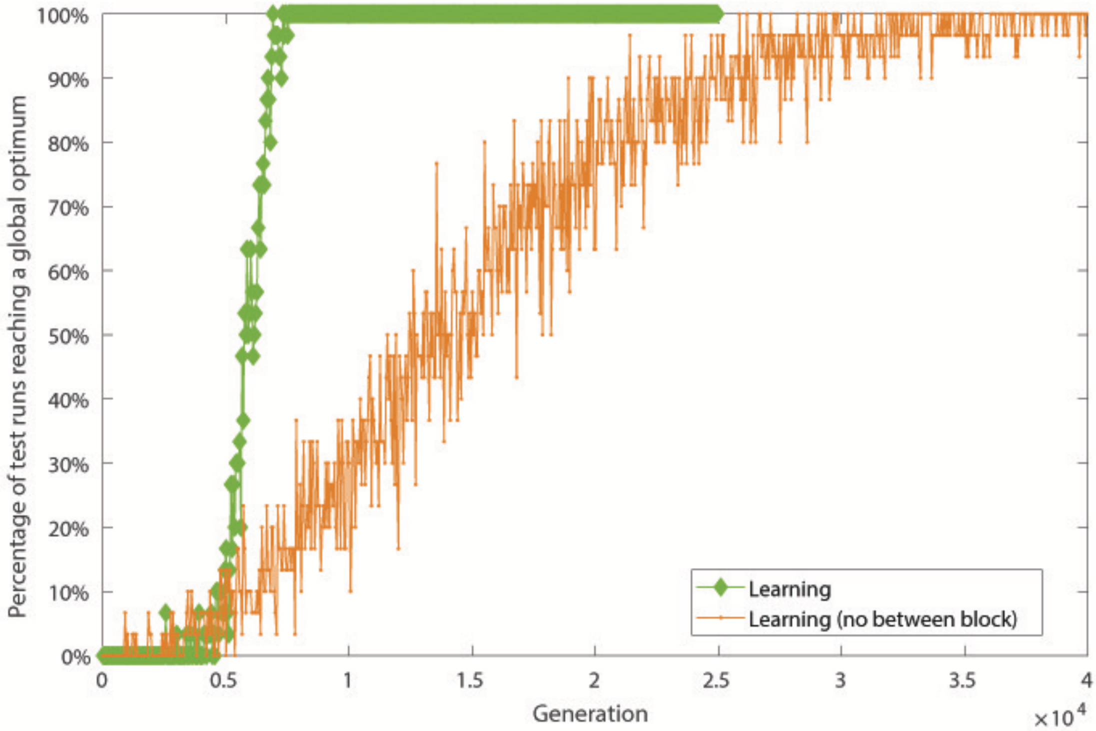


FIGURE 6.8: The performance of learning with and without the evolution of between-blocks is plotted every 500 generations for the Correlated Behaviours model using the MC problem. The light green line with diamond markers shows standard learning (with evolution of the within and between-block correlations). The darker green line with dots shows evolution of within-block correlations only. This demonstrates that whilst evolution is quicker when between-blocks are evolved, it is not a necessary condition for this model to find a globally optimal solution.

As the evolution without the between-block correlations takes longer, this would suggest that, whilst not critical to reaching the global optimum, these between-block correlations do have a role in accelerating the progress at which optimal fitness phenotypes can be achieved. It is worth noting that the between-block correlations offer learning the potential to discover global optima at another level of hierarchy. This is not required for the MC problem because it only has two levels encoded by the strong and weak constraints. There is potential to explore further rescaling using different problem domains (see Future Directions in Chapter 8).

6.4.4 Local Connections can be Resolved by Learning

To verify that the Correlated Behaviours model is not reliant on an explicit block structure of the MC problem, the performance of the system is assayed on the Concentric Squares problem, which defines the fitness landscape in terms of localised connections (as described in Section 4.1.2). Figure 6.9 shows that, in keeping with the result using the MC problem, the learning case quickly finds the global optimum, whilst for the no-learning case, there is little appreciable increase in fitness over the evolutionary timescale tested.

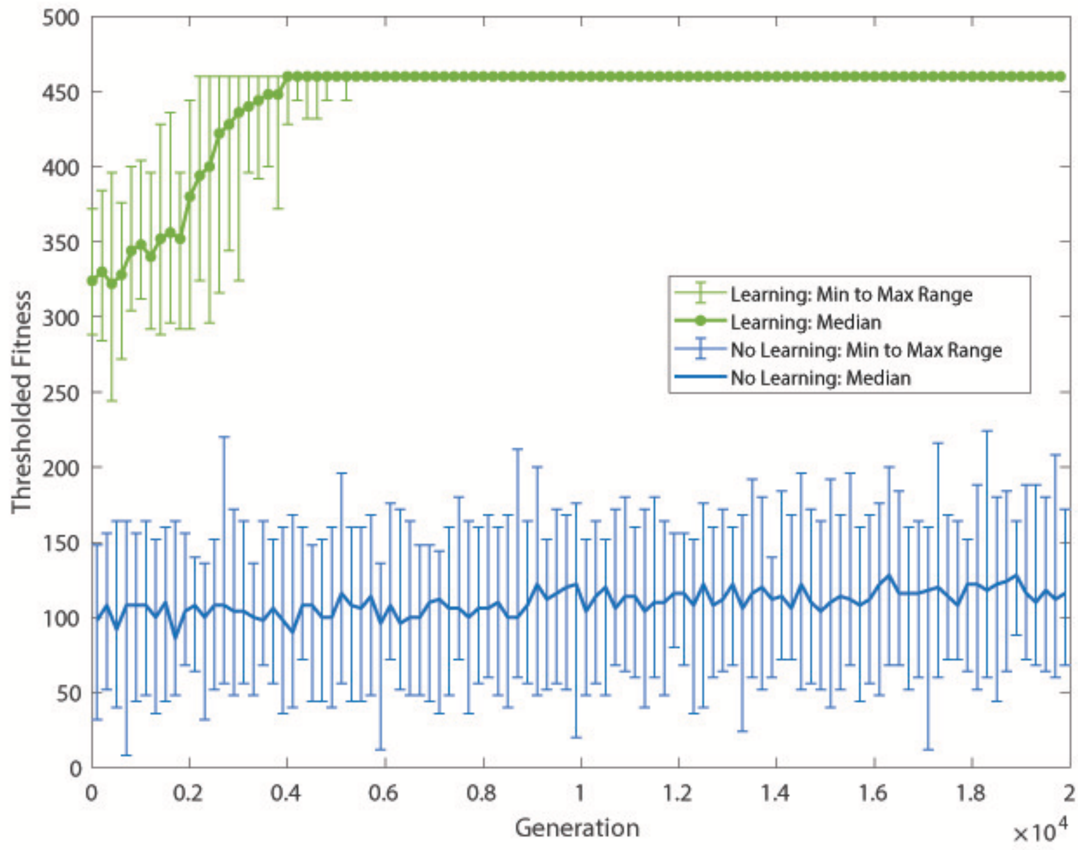


FIGURE 6.9: The performance of learning in comparison to no learning is assayed for the Concentric Squares problem. In common with the MC problem, learning quickly finds the global optimum (green line with circle markers) whereas the no learning case does not see an appreciable increase in fitness (blue line).

It should be noted, that the local connections are a function of the fitness landscape, and in common with the ESDP model (as shown in Figure 5.8) all connections in the correlation matrix are evolved.

6.4.5 Global Optima are Reliably Discovered With Sparse Connections

In common with the other models in this thesis, there is potential for all innate behaviours to correlate with all other behaviours. Whilst there is a high degree of pleiotropy to many traits [129], it would be unusual for all behavioural traits to be genetically correlated to all others - this is the result for both the MC (as seen in Figure 6.7 (V)) and CS problems (not shown). Indeed, given that any element of the correlation matrix (B) is subject to mutation, this model facilitates connections being made between all traits. Applying a parsimony pressure on connection strength by using a 'cost of connection' (as per Kouvaris et al. [59]) or limiting the number of connections that can be made is likely to be a more biologically realistic scenario. This is briefly assayed by setting at random which connections are subject to mutation and therefore which innate behaviours can be connected to which others. To test this, the correlation (B_{ij}) is subject to mutation if the connection matrix has a value below a threshold for allowing a connection: $F_{ij} \leq q$ where, q is a connection density parameter in the range zero to one specified at the beginning of evolution, and F is a matrix of the same size as the B matrix, populated with uniform random values between zero and one:

$$F = (f_{ij}) \in \mathbb{R}^{n \times n}$$

$$F_{ij} \sim U(0,1), \forall i, j$$

For a range of rates of connection $0.04 \leq q \leq 1$, the global optimum is consistently reached. This is demonstrated in Figure 6.10 which plots the minimum, maximum and average number of mutations[5] for evolution and learning to find a globally optimal phenotype for thirty runs, where consistently reaching the global optimum is defined to be if a phenotype with globally optimal fitness is expressed for 20 consecutive generations. As can be seen, the average number of mutations to reach the global optimum fitness required is broadly consistent for all connection densities but becomes less consistent where the connection density is below 0.06, i.e. where one locus is connected with an average of six others. Where the connection density q is 0.02, the global optimum is not reliably discovered within the 200,000 generation limit.

[5]For convenience, the average number of mutations is calculated by multiplying the number of generations by the number of mutations per generation (which in this model is n) and by the probability that each mutation is allowed (which equates to q)

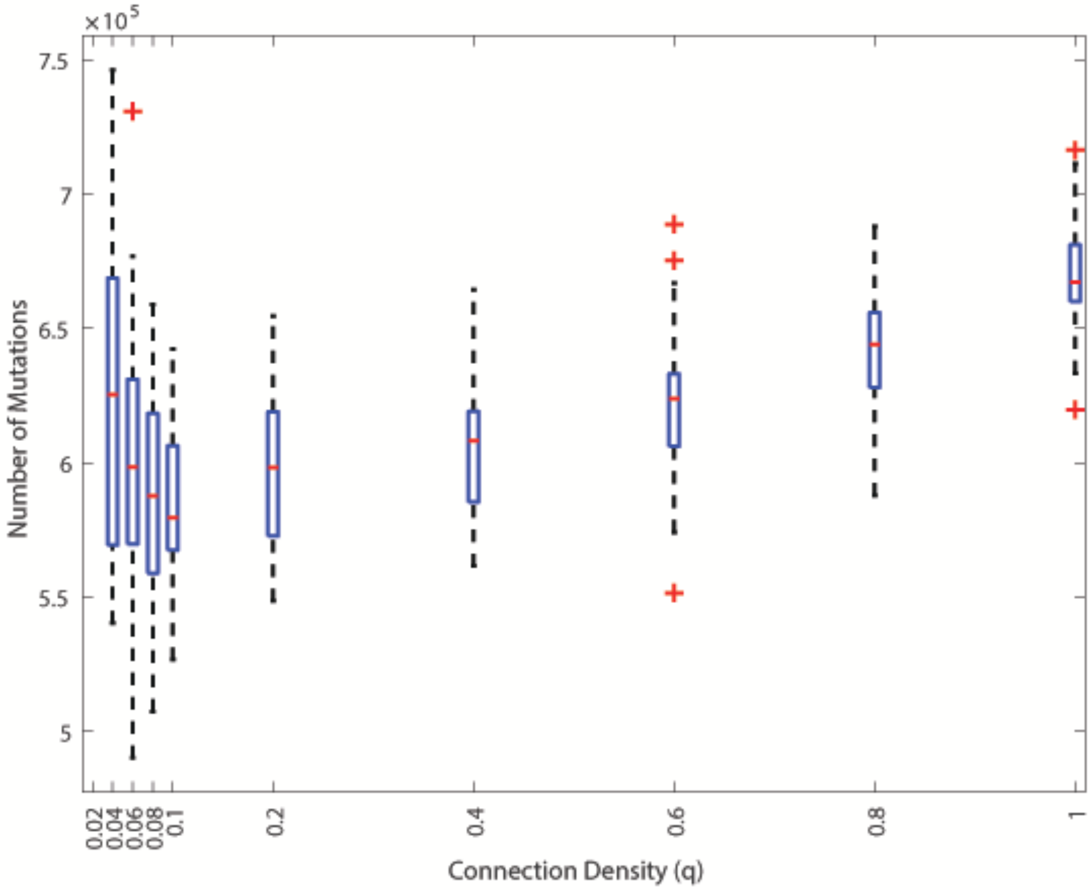


FIGURE 6.10: Box plot of the number of mutations required to reach the global optimum fitness consistently for 20 generations for 100% of thirty runs for the range of connection densities $q \in \{0.04, 0.06, 0.08, 0.10.2, 0.4, 0.6, 0.8, 1.0\}$. The experiment is limited to 200,000 generations.

However, as the connection rate tends towards zero, there is an increasing probability for a trait to become completely disassociated from all other traits[6]. When this happens the traits become independent so that some traits are only dictated by the environmental input - they cannot be changed by learning or innate behavioural expression and so the global optimum phenotypic configuration is not consistently reached.

This result suggests that although assimilation of the learning is reliant on canalisation due to pleiotropic effects, a high degree of pleiotropy is not a condition of assimilation.

[6]There is likely to be an intermediate stage when traits become co-dependent through a common connection with a third trait.

6.5 Analysis and Discussion

The results presented in this chapter show that a generative model of learning significantly accelerates finding a globally optimum phenotype. In addition, as learning scales-up from changing single behaviours to coordinating multiple behaviours the rate of evolutionary change increases. Moreover, the Baldwin Optimizing Effect is observed: evolution with learning reliably discovers the global optimum phenotypes whereas without learning, locally optimal phenotypes are discovered - which on occasion are globally optimal - but this discovery of local optima is not sustained between generations and is random with respect to the environmental input

6.5.1 Have the Conditions for Assimilation Been Relaxed?

In Chapter 3, the necessary conditions for assimilation of learning were assessed - candidates being (1) a cost of learning, (2) a correlation between movement in genotype space and phenotype space and (3) a stable learnt outcome for the genotype to assimilate to. Whilst the conditions are not clear cut - for example, drift can also enable attainment of a phenotype with globally optimal fitness, it is worth reviewing the Correlated Behaviours model in the context of each of these conditions to understand to what extent these conditions have been relaxed.

6.5.1.1 Cost of Learning

Where learning alters the phenotype's traits after expression of the genetically determined traits, there is potential for the mutational effect to be masked by that learning: mutations do not make a fitness difference because learning provides the same fitness for the mutated and unmutated phenotype (flattening the fitness landscape as shown in Figure 3.1 of Chapter 2). In non-epistatic, trait-based models of learning and evolution, a cost of learning is often deployed to reapply a selective gradient towards the adaptive peak.

There is potential for a similar 'hiding effect' to occur in this Correlated Behaviours model: small changes to the genome caused by mutations are very likely to lead to the production of the same phenotype[7]. Therefore, if learning is considered on its own, there is no differential fitness signal for evolution to act on.

[7] The mutation to the matrix of innate behaviour correlations (B) is relatively small which makes it unlikely that the learnt phenotype resulting from the mutated correlation (for convenience lets call this l') is different to the phenotype (l) resulting from the unmutated correlation matrix. The is because the probability that a random threshold r not selecting the same set of traits for a given learning step is small - for a the mutated loci $B_{i,j}$, and mutation size γ, r would need to be in the range $b_{i,j} \leq r \leq (b_{i,j} + \gamma)$. Even if one of the learning steps were in this range, the subsequent learning steps are likely to change the sign of multiple traits and so, the fitness difference would be swamped by the fitness effects of the other trait flips.

In addition, a cost of learning similar to that implemented by Hinton & Nowlan [49] is not applicable because in the scenario modelled, learning is open-ended and there are always assumed to be the same number of learning steps[8]. Even if a cost of learning were applied, it would not transmit a correct signal to selection as to the direction of the adaptive peak. If learning ceased when a local or global optimum was encountered by the learnt phenotype and a cost of learning based on the number of learning trials applied (as per Hinton & Nowlan [49]), there would still be no adaptive signal. The number of learning steps would be highly stochastic given that the number of traits flipped in a single learning step is dependent on a random threshold and a random focal bit. Therefore, the number of learning steps would not provide a good fitness signal to selection.

However, in this model, there is no hiding effect observed and no need for a cost of learning - the expression of the magnitude of the innate behaviours during learning effectively reapplies the fitness difference created by the mutation at each learning trial. In this case, because assessment of the fitness of the innate behaviours is achieved by applying the connection strengths in the correlation matrix (B) during the fitness assessment of each learning trial (as per equation 6.6), a fitness difference will be maintained between the mutated and unmutated genotype (B and B') at each step. At the end of learning this fitness is used to assess the adaptiveness of the mutation and so the fitness difference provided by the expression of innate behaviours enables evolution.

6.5.1.2 Neighbourhood Correlation

As discussed in Chapter 3, neighbourhood correlation is defined by Mayley [74] in terms of the size of movement in genotype spaces as compared to movement in phenotype space. We found that even when considering size of movement, the concept of neighbourhood correlation was not clear-cut because it was dependent on several underlying assumptions, most notably the fitness effect of a single learning step as compared to the fitness effect of mutations.

This Correlated Behaviours model does not have a neighbourhood correlation issue: although learning causes large movements in phenotype space, a mutation can make a fitness difference that sends the correct fitness signal to selection. Adaptive mutations incrementally strengthen correlations which eventually allows the generation of globally optimal phenotypes. The genome's capacity to accumulate the fitness signals from learning allows the genome to slowly move towards an optimal phenotype. Consequently, there are no large learning steps that cannot be eventually be replaced by mutation as was the case in the Mayley model.

[8] This assumption is based on the premise that the individual will not know if the next learning step - testing the innate behaviours in the next environment - will be better or worse.

In addition, when discussing neighbourhood correlation Mayley [74] also describes another facet of neighbourhood correlation where the genotype is at one adaptive peak and the learnt phenotype is at another adaptive peak. In this model, the phenotype before learning is highly likely to be in a different basin of attraction to phenotype after learning. This does not cause issue because the expression of the innate behaviours means that the fitness effect of the mutation is judged in the context of the learnt phenotype, not the genotype.

6.5.1.3 Stable Learning Outcome

Unlike a one-to-one mapping between genotype space and phenotype space, a correlation model is capable of encapsulating and generalising regularities in the fitness landscape [59, 126]. Kashtan, Noor & Alon [53] argue that even though the environment changes both temporally and spatially, these changes have consistent sub-goals. They use the behaviours of *"feeding, mating and moving"* as examples of consistency required across changing environments. Their simulations showed that modular varying goals (MVG) - where the MVG problem structure was decomposable into sub-goals that were shared across environments - increased the rate of evolution. In this Correlated Behaviours model, the ability to encapsulate regularities from the environmental inputs means that as long as the fitness signal provided by learning provides a better than random adaptive signal about the fitness landscape in general, evolution will on average change the correlation strengths in a way that captures those regularities. Indeed, without a regularly changing environment, learning is likely to ascend to the same local optimum at every generation and evolution will canalise the genotype to one that only generates phenotypes at that optimum.

We can therefore conclude that, in a canalisation model of genetic assimilation, due to the way that repeated exposure to different environmental inputs allows the genotype to gradually canalise to the average of those environments, a stable learning outcome is an inhibitor to the discovery of a globally optimum phenotype a rather than a necessary condition for assimilation.

6.5.2 Could this Model Apply to Physical Traits?

As discussed in Section 2.1, both innate behaviours and developmental plasticity have multiple characteristics in common; both are usually environmentally sensitive, change at the population level and are more genetically determined than learning. So could the structure of the model presented in this chapter also serve as a model of learning's effect on the evolution of physical traits?

For physical traits, one can assume that the expression of morphological characteristics precedes learning, whilst for innate behaviours, as is the case with this model, the assumption is that learning alters the expression of innate behaviours: learning and the expression of innate behaviours are happening in parallel. Consequently, the method of getting a fitness difference between the mutated and unmutated versions of the population mean phenotype, via the consequential expression of innate behaviour, is not applicable to physical traits. Additionally, developed physical traits are usually thought of as irreversible within a lifetime and therefore not usually altered through the learning process as are innate behaviours[9]. Fundamentally, the relationship between morphological traits and evolution is different to that between learning and innate behaviours and this model would require major revision to be an appropriate simulation of the interaction between learning and the evolution of physical traits.

Since physical traits are an important constraint to behaviour, a model of the interaction between the evolution of physical traits and learning - especially with regard to West-Eberhard's hypothesis that learning will correlate the evolution of physical traits (as discussed in Section 2.7.4), remains a challenging area ripe for future investigation.

6.6 Conclusion

The Correlated Behaviour model confirms that learning which finds good solutions in trait space can be genetically assimilated through canalisation of the genotype. Although a like-for-like comparison of this model and the ESDP model is problematic, the Correlated Behaviour model reliably reaches a globally optimal fitness with significantly fewer evolutionary fitness evaluations than the ESDP model.

Importantly, the generative action of learning, constrained by innate behaviours means that with learning globally optimal phenotypes are reliably found that are not without learning. This demonstrates a form of Baldwin Optimizing Effect.

In this model, learning appears to be much more powerful in determining evolutionary outcomes than the environmental input and expression of innate behaviours, because the rate of evolution with learning is significantly faster than without. This is supported by the result that shows evolution is not significantly faster where there is evolution of between-block correlations as compared to without evolution of between-block correlations. In the next chapter, we will explore an indirect learning mechanism and see if this has a similar effect on the speed of reaching high fitness phenotypes.

[9]There are instances where learning alters physical traits such as the morphological changes that occurred when the two legged goat learnt to walk on its hind-legs [129].

Chapter 7

Environment Selection Model

7.1 Introduction

In common with the Correlated Behaviours model presented in the previous chapter, this third and final model also considers how learning and innate behaviours can positively influence each other to change the rate and trajectory of evolution, but this time through an indirect causal route.

The model presented here examines a two-way feedback mechanism where learning alters the selection pressures on genetically determined behaviours by changing the environment that the phenotype experiences thereby impacting multiple behavioural traits simultaneously. In turn, the innate behaviours expressed in a given environment bias what is learnt. Unlike the Correlated Behaviours model, learning does not directly specify individual behavioural traits. Consequently, in this simulation, there is no direct linkage between learnt and innate behaviours and the indirect mechanism modelled allows both learning and innate behaviours to influence the behavioural expression without competing for control of the behavioural phenotype. This is in contrast to prior models of learning and evolution (reviewed in Chapter 2).

The interactions between learning and evolution simulated in this model are based on the interactions summarised in Figure 2.4 (C); this original flow has been augmented by making explicit the interaction between genetic variation and innate behaviours and between learnt behaviours and the environment, as shown in Figure 7.1.

A critical assumption in this model is the assertion that, as discussed in Section 2.1.1, innate behaviours are plastically expressed based on environmental input; for example, the innate behaviour of a goose rolling an egg back into its nest is triggered by the goose seeing an object that looks somewhat similar to an egg [44]. In this way, learning can influence the expression of innate behaviours based on changing the environmental input to those innate behaviours.

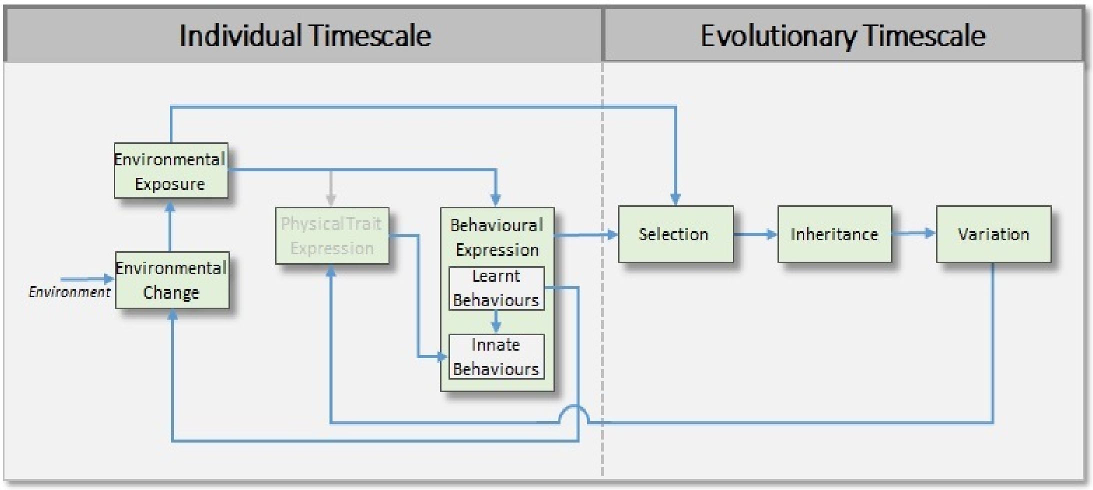


FIGURE 7.1: In the Environment Selection Model, behavioural expression is triggered by environmental stimuli and is a combination of innate and learnt behaviours. Trial-and-error learning allows the individual to select the best environment given its expression of innate behaviours in each trial environment; this effectively changes the environment in which its fitness is determined and therefore selection acts. Genetic variation alters the expression of innate behaviours in the environment tested by learning. Physical trait expression is not considered in this model and therefore the text in that box is greyed in this figure, as compared to Figure 2.4 (C).

Whilst in this experimental set-up, the environment experienced by the phenotype (the effective environment) changes during an individual's lifetime, this model is not intended to be a simulation of niche construction: the organism does not make a lasting impact on the environment and consequently there is no ecological inheritance [60]. Therefore, unlike Figure 2.4 (C), the *'Environmental Change'* box is placed firmly within the individual timescale area of the diagram, reflecting the temporary nature of the environmental variability.

The behaviour inspiring this model is that of habitat selection which is thought to be one of the most basic behavioural choices that an organism can make [47]. In this case, the individual's genotype controls the innate behaviours expressed by the phenotype based on environmental input. The organism is also capable of within-lifetime learning allowing it to choose, through trial-and-error, in which habitat it is best to reside. Each new generation learns the best habitat from a random selection of potential habitats. The individual then stays in their chosen environment for the majority of their life and, since innate behaviours have varying fitnesses depending on the environment, this choice of lifetime environment provided by learning will have a significant fitness impact. The learnt choice of environment is determined by the best fitness attained by the phenotype's innate behaviours tested in each of the environments. A simplistic example of this scenario is learning being able to determine which locations provide the best fitness improving food[1] using innate foraging behaviours. For each generation, the

[1]Based on an organism's having an adaptive learning feedback mechanisms - good and bad context-specific sensations that can be remembered - being specific to motivations that are tuned to survival [129].

fitness of the individual for the purposes of selection is therefore based on the fitness of the innate behaviours expressed in the habitat chosen by learning.

As will be shown from the results of the simulations, the mutual, indirect feedback mechanism between learning and innate behaviours allows learning - as an adaptive form of plasticity - to accelerate the rate of evolution of innate behaviours, but in this case, do not achieve evolutionary outcomes that cannot be achieved through evolution without learning. It will also be shown that as evolution progresses, the range of phenotypes that are produced narrows, suggesting some degree of genetic assimilation of the learnt behaviours (whether true genetic assimilation is achieved or not is discussed in Section 7.5.1).

7.2 Model Structure

The Environment Selection model encodes learning and innate behaviours in the following way: the expression of the innate behaviours in a given environment is determined by taking the environmental input and through an iterative process, expressing the behaviours for that environment based on the genetic correlations between innate behaviours. The individual, therefore, has a fitness in any given environment based on the combination of innate behaviours expressed given an environmental input. Genetic variation is introduced through mutation to the genetic correlations and therefore it is the connections between innate behaviours that are inherited by the next, non-overlapping generation.

The Modular Constraint (MC) problem is used as a representation of hierarchical behaviours where good combinations of low-level behaviours are deemed to have a fitness that is a local optimum and where fit combinations of higher-order behaviours have optimal fitness. Therefore, for combinations of behaviours to be maximally fit, the right combination of both low-level and high-level behaviours is required (see Section 4.2.2 for a more detailed discussion of the applicability of the MC problem to learning). We follow previous models by assuming the fitness consequences of sets of low-level behaviours are strongly correlated, whereas the correlations between high-order behaviours are weak. Further in this particular schema, all correlations between behaviours are genetically determined: learning cannot directly affect which behaviours are correlated.

Learning is introduced by allowing the organism to select the best environment in which to reside through sequential trial-and-error testing of a defined number of environments. The chosen environment being that in which the innate behaviours are fittest, based on the environmental input to those behaviours. The set of environments that are explored through learning changes at each generation. This is therefore a 'dumb' model of operant conditioning; learning provides the organism with the ability to explore different environments and make a choice based on the overall match of innate behaviours to that environment. This structure provides a simple mechanism for genetically determined innate behaviours to be modified by trial-and-error learning via the environment. In this way, learning can change the selection pressures on the genetically prescribed behaviours without competing directly for control of the behavioural phenotype.

As a simplifying assumption, the influence of the plastic expression of physical traits is not considered, and therefore it is assumed that the physical machinery of learning and innate behaviour expression does not improve over the evolutionary timescale.

Unlike most models of learning, including the classic Hinton & Nolan [49] model and its many derivatives, we do not assume that learning stops when the individual reaches an adaptive peak. In this model, each individual will continue to test the same number of environments at each generation. The rationale for this is that, whilst the individual can sense in which environments their behaviours are fitter, they cannot know if the next environment will be better.

An important assumption underpinning this model is that once a defined number of potential environments have been tested, the individual will choose the best habitat in which to reside for the majority of their life-cycle. Consequently, their lifetime fitness for the purposes of selection is dictated by the chosen environment: time spent in other environments is not taken into account. This assumption is common to most models of learning and evolution, where the learnt phenotype's fitness is assumed to be the fitness after all learning trials have been completed rather than the average fitness over each of the learning trials.

With this model, we again use the 'Strong Selection, Weak Mutation' (SSWM) assumption [41], where the population can be represented by the population mean genotype and resultant phenotype, as discussed for the ESDP model. This simplification is needed primarily due to the computation time required to simulate behavioural choices: modelling a significant population is impractical. However, as we are also modelling in-lifetime behaviours, this adds the assumption that the best set of behaviours is expressed by the population mean genotype. Since lifetime learning is lost after each generation and the learning outcome is also not consistent across generations, a population is unlikely to materially alter the dynamics of the simulation. As is the case with the other models constructed in support of this thesis, only correlations between genetically determined behaviours are modelled rather than the specific trait values of individual genetic loci.

7.3 Model Detail

In common with the Correlated Behaviours model, the Environmental Selection model uses a matrix of genetic correlations (B) where this matrix represents correlations between innate behaviours. This matrix is used to determine how innate behaviours are deterministically expressed (see detail below). The specification of the environment input vector, correlation matrix (B), mutation matrix (J), fitness function (w) and fitness thresholding function (θ) are the same as that for the ESDP and Correlated Behaviours models, provided by equations 5.4, 5.1, 5.2, 5.6 and 5.7 respectively.

The detailed relationship between model functions is shown in Figure 7.2 and is further defined in the following sections.

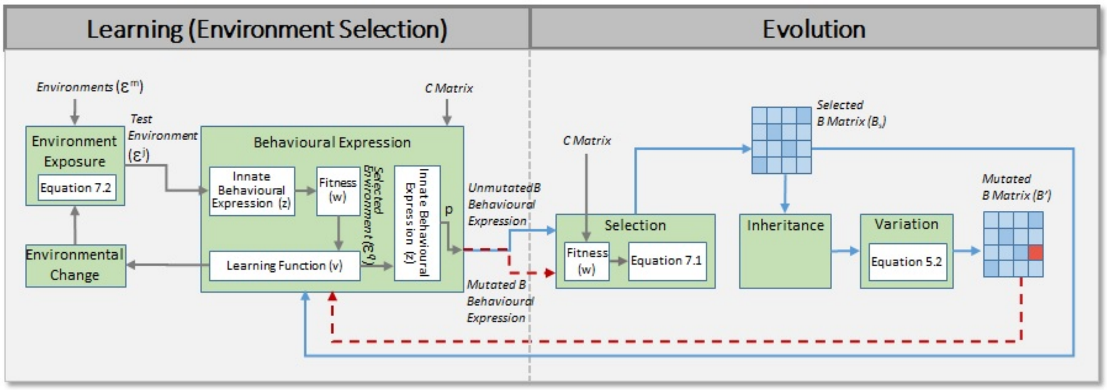


FIGURE 7.2: Following the structure shown in Figure 7.1, an illustration of the interaction between Environmental Selection model functions. Under the strong selection, weak mutation (SSWM) assumption, the learning function (z) tests the fitness of behavioural expression (using functions u and w) in a given test environment ($\boldsymbol{\epsilon}^j$) by exploring the set of candidate environments ($\boldsymbol{\epsilon}^m$). Once the best environment ($\boldsymbol{\epsilon}^q$) is chosen by learning, the behavioural expression in this environment (using function z) becomes the lifetime behavioural phenotype ($\mathbf{p}$). On the evolutionary timescale, selection tests the lifetime behavioural expression of both the unmutated (B) and mutated (B') correlations matrix. The mutation to the B matrix is generated using equation 5.2 and equation 7.1 is used to select whether the mutated or unmutated matrix is inherited to become the next generation. At each generation, a new, random set of candidate environments is generated using equation 7.2. The flow of the unmutated form of the correlation matrix (B) and resultant developed phenotype is denoted by solid blue lines whereas the flow of the mutated correlation form of the correlation matrix (B') and derived developed phenotype is represented by red dashed lines. Again, in common with the ESDP model, the selective environment is defined by the C matrix which varies dependent on problem type.

The model simulating the interaction between learning and evolution of innate behaviour is based on a vector representation of phenotypic behaviours being the product of a matrix of genetic correlations interacting through an evolution function, a learning function (v), a behavioural expression function (z), and a fitness function (w). The population on which selection acts is represented by the population mean phenotype that has a set of innate behaviours defined by a vector of n real values where:

$$\mathbf{p} = p_1, p_2, \ldots, p_n \in \mathbb{R}^n$$

7.3.1 Evolution

As per the models presented previously in this thesis, the correlations between innate behaviours are represented by the correlation matrix B where b_{ij} represents the connection strength between the gene controlling innate behaviour i and the gene controlling innate behaviour j. Mutations to elements in the B matrix, therefore, represent variations in the population mean connection strength between innate behaviours.

As was the case for the ESDP model but not the Correlated Behaviour model, at the start of evolution, the B matrix is initialised to zero, denoting a complete absence of correlations between behavioural traits.

$$b_{ij} = 0, \forall i, j$$

The mutation of the correlation matrix is specified using equation 7.1:

$$B_{s+1} = \begin{array}{rl} B_s + \gamma J^{(kl)}, & \text{if } w(z(B_s + \gamma J^{(kl)}, v(B_s + \gamma J^{(kl)}, \boldsymbol{\epsilon}^m))) \geq w(z(B_s, v(B_s, \boldsymbol{\epsilon}^m))) \\ B_s, & \text{otherwise} \end{array} \tag{7.1}$$

As stated previously, the fitness function (w) is the same as that used in the ESDP model and thus is as per equation 5.6. The behavioural expression is controlled by the behavioural expression function z that develops a behavioural phenotype based on the genetic correlations (in matrix B) and the environmental inputs $\boldsymbol{\epsilon}^q$ and is defined by equation 7.4. The environmental input $\boldsymbol{\epsilon}^q$ is selected by the learning function v defined below. The sequence of mutations for that generation completes at $B_{(s=n^2)}$, so for all experiments where the B matrix is evolved, the number of mutations of B per generation is set as the square of the number of genes in the genotype. The magnitude and sign of the mutation is defined by γ, which is a random value drawn uniformly from the range $\pm d$. Unless otherwise stated the value of d was set to be consistent with the ESDP model at 2×10^{-5}.

7.3.2 Learning (v function)

The learning function v chooses the best environment $\boldsymbol{\epsilon}^q$ from a set of candidate environments $\boldsymbol{\epsilon}^m$ (as described below).

Since learning is not inherited, it takes place after every test on the mutation of B. During the lifetime learning cycle, a set of m candidate environments are tested with the best environment deemed to be that where the population will reside. The set of m test environments changes at the end of every generation (not after every test mutation of B). Each candidate environment is defined by a randomly generated environment vector with n continuous values drawn uniformly between -1 and 1. The collection of random environment vectors is defined by:

$$\boldsymbol{\epsilon}^m = (\boldsymbol{\epsilon}_1^m, \boldsymbol{\epsilon}_2^m, \ldots, \boldsymbol{\epsilon}_n^m), \boldsymbol{\epsilon}_i^j \sim U([-1,1]), i \in \{1,2,\ldots,n\}, j \in \{1,2,\ldots,m\} \tag{7.2}$$

Therefore, the number of elements of the environment matches the number of traits in the phenotype. Unless otherwise stated, a set of twenty environments are tested and this set changes at each generation (section 7.5 explores the effect of the number of environments on the speed of evolution).

The best environment $\boldsymbol{\epsilon}^q$ is selected using the learning function v which determines which environment returns the fittest behavioural phenotype based on the innate behaviours exhibited in that environment, so that:

$$\boldsymbol{\epsilon}^q = v(B_s, \boldsymbol{\epsilon}^m), q = \underset{j \in \{1,2,\ldots,m\}}{\operatorname{argmax}} \ w(z(B_s, \boldsymbol{\epsilon}^j)) \tag{7.3}$$

7.3.3 Innate Behavioural Expression (z function)

The innate behaviour expression function z uses a correlations matrix B to derive the set of behaviours in the test environment $\boldsymbol{\epsilon}$, in the following way. At time t, the behavioural phenotype is defined by $\mathbf{p}_t$ using the behaviours expression function z. Behavioural expression starts with the environmental input being tested by learning $\boldsymbol{\epsilon}$, and therefore $\mathbf{p}_{t=0} = \boldsymbol{\epsilon}$. The resulting set of innate behaviours is defined by iterating for η steps, a behavioural development function similar to that of the ESDP model's development function, which determines the expressed behaviours based on the genetic correlations and the environment:

$$\mathbf{p}_{t+1} = (1-\tau)\mathbf{p}_t + \sigma(\mathbf{p}_t B) \tag{7.4}$$

Where τ is a decay rate that helps contain the overall values in the $\mathbf{p}$ vector. In common with the ESDP model, for all experiments in this chapter $\tau = 0.2$ and the number of steps η is 20, and the same σ function ($\sigma(x) = tanh(x)$) is applied to all elements of the $\mathbf{p}$.

For experiments where there is no in-life learning, the learning steps still occur, but the choice of environments is restricted to the first random location ($m = 1$). Effectively, the fitness is the innate behavioural expression in one random environment. So, unlike the Behavioural Correlations model, there is no random exploration of the phenotype space when there is no learning. However, whilst there is no in-life testing of other habitats for the no-learning case, it is also assumed there is environmental variation between generations.

7.3.4 Behavioural Phenotype's Fitness (w function)

The fitness of the phenotype's combination of behaviours in the best environment at the end of development $\mathbf{p}_{t=\eta}$ is assessed by the phenotype's closeness to the target phenotypic behaviours using the fitness function w, as per equation 5.6. As explained for the ESDP model, since it is the sign of trait that deems if it matches the target rather than the magnitude when comparing experimental results, the thresholded fitness function θ is used.

For the Environment Selection model, the MC problem with a p value of 0.01 is used as the fitness landscape as described in Section 4.1.1. However, due to the extra computational requirements that the modelling of lifetime learning requires, the size of the constraints matrix is changed to being ten sets of highly correlated behaviours with each set being six individual behaviours. Keeping the same number of blocks ensures the probability of encountering one of the global optima, given a uniform distribution of local optima, is broadly the same as the other models presented in this thesis.

7.4 Results

This section outlines the key results from the analysis of the performance of the Environment Selection model.

7.4.1 Learning Accelerates the Rate of Evolution

First, we test to see if the indirect learning mechanism accelerates the rate of evolution in comparison to where no learning is present. Figure 7.3 shows that both with and without learning, the fitness of the phenotype representing the population-mean fitness rises as evolution progresses. This would suggest that even though there is a random starting environment, evolution can find a combination of behaviour that is fit for the environment. However, with learning, fitness rises significantly faster than without learning suggesting that evolution is quicker when the population can choose its environment through trial-and-error learning.

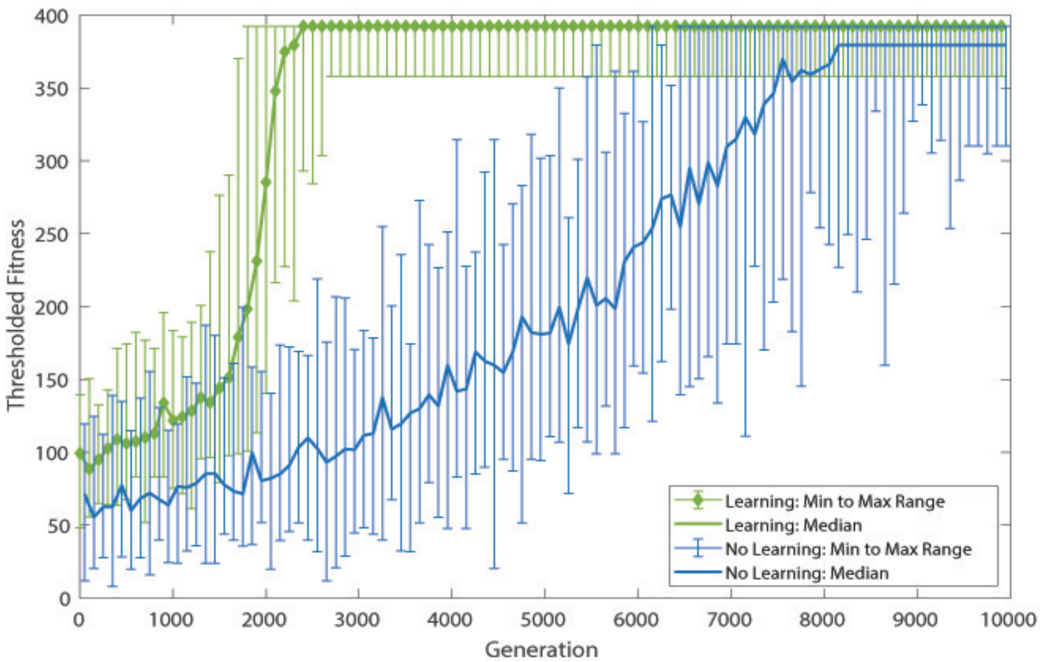


FIGURE 7.3: Comparison of fitness of combinations of innate behaviours with and without learning for the Modular Constraints problem. The threshold fitness of the population-mean phenotype, under the SSWM assumption, is plotted at every hundredth generation. The error bars show the minimum to maximum range of fitnesses achieved at that generation over 30 runs using the same parameters. The darker line shows the median value over the same 30 runs. With learning, shown in green with diamond markers and without learning, shown in blue (no markers).

To check that the performance of the Environment Selection model is not specific to the structure of the MC problem, we also test evolution and learning using the Concentric

Squares problem. This CS problem does not have the same block structure as the MC problem[2] as it only contains local constraints.

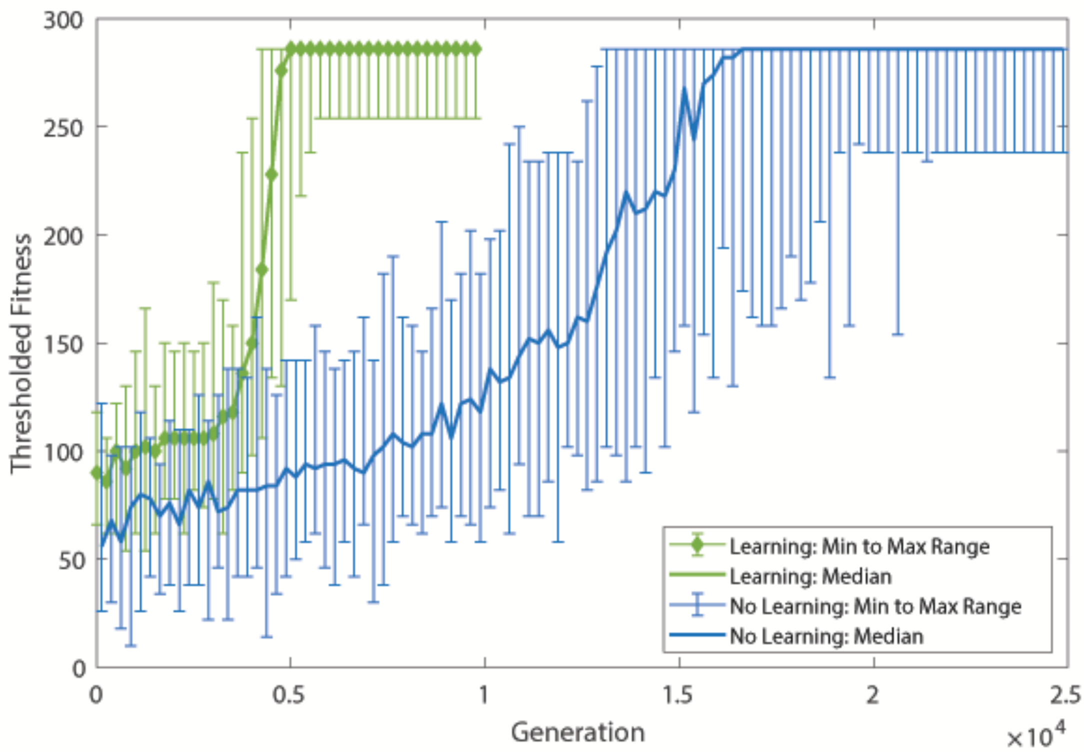


FIGURE 7.4: Comparison of fitness of combinations of innate behaviours with and without learning for the Concentric Squares problem. The threshold fitness of the population-mean phenotype. The error bars show the minimum to maximum range of fitnesses achieved at that generation over 30 runs using the same parameters. The darker line shows the median value over the same 30 runs. The data is plotted at every 250 generations, with the second plot offset by 100 generations. With learning, shown in green with diamond markers and without learning, shown in blue (no markers).

As Figure 7.4 shows, the performance with and without learning is similar to that for the Modular Constraints problem (as shown in Figure 7.3): with learning quickly finding the global optimum. Without learning, the global optimum is found less quickly when using the same mutation rate as with learning (in this case $d = 1 \times 10^{-5}$).

The Environment Selection model therefore clearly demonstrates a Baldwin Expediting Effect for both test problems.

7.4.2 Speed of Evolution is Sensitive to Number of Learning Trials

We now test whether the extent of learning (the number of environments tested) changes the rate of evolution. The results presented previously are for trial-and-error learning,

[2]The short range connections are defined in terms of traits that are adjacent when viewing the genotype as divided into rows of a defined size and therefore there is an element of building-block structure (i.e. the rows)

selecting the best environment from a set of twenty random environments. It should be expected that reducing the number of environments would reduce the impact of learning, therefore, slow the speed at which evolution reaches optimal fitness, to the point where only testing one environment is the no-learning case.

The results for learning based on the learning exploring 1 (no learning), 3, 5, 10, 15, 20 and 25 random environments are compared in Figure 7.5. Reduction in the number of environments tested by learning reduces the expediting effect of learning, but not linearly (halving the number of environments tested does not half the rate of evolution).

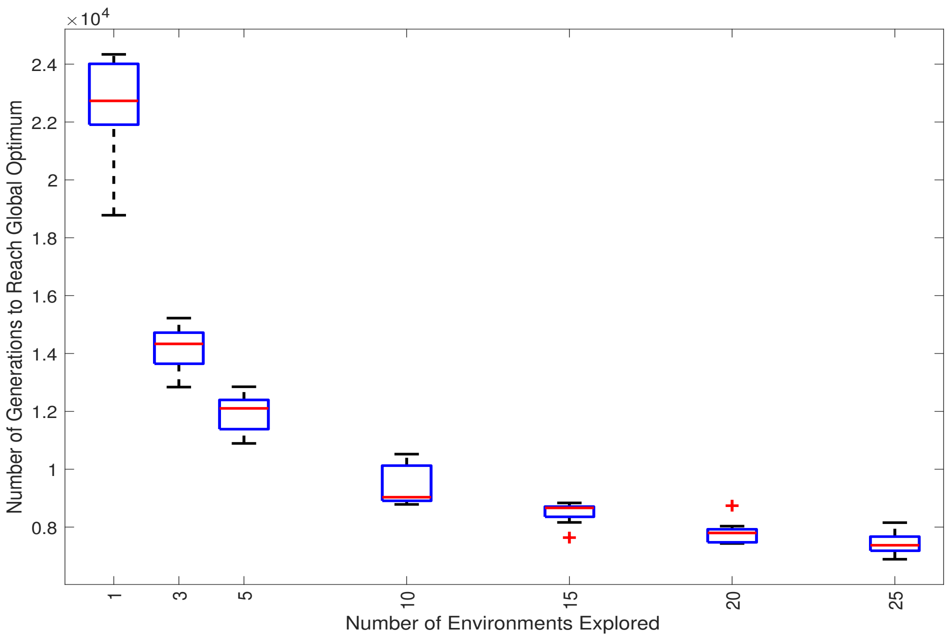


FIGURE 7.5: Comparison of the effect on the rate of evolution due to the number of environments explored by trial-and-error learning. The minimum and maximum number of generations to reach a global optimum is shown by the bars. The red line shows the median value over the same eight runs.

It should be noted that the correlation matrix mutation parameter (d) is lower than previous experiments at 5×10^{-6}. This was primarily due to the requirement for all tests to reach a global optimum, including the no learning case.

When considering the simulation from a pure computation perspective, there are approximately twenty times the number of fitness evaluations with learning than without learning when twenty environments are explored. Therefore, learning is not offering a computational benefit; this is common to the original H&N model and other models of learning and evolution: if a mutuational step and a learning step are considered as equal then learning is offering no optimization benefit. This supports the intuition that, for this model, it is the difference in relative timescales between learning and evolution that enables and Baldwin Expediting effect rather than any intrinsic optimisation due to the action of learning. Learning acts like evolution finding fit phenotypes using fitness evaluations - it has a fitness screening effect [129].

7.4.3 Learning Finds Optimal Combinations of Innate Behaviours More Reliably

As discussed in Chapter 2, as well as the Baldwin Expediting effect, where learning accelerates evolution, there is potential for a Baldwin Optimizing Effect where learning finds optima that would not normally be found. Figure 7.6 shows the number of times that a global optimum is encountered with and without learning using the standard model parameters. The no-learning case shown in 7.6 (blue line) suggests that whilst learning more reliably reaches the global optimum fitness, without learning is also capable of reaching the global optimum more than it would do by chance[3].

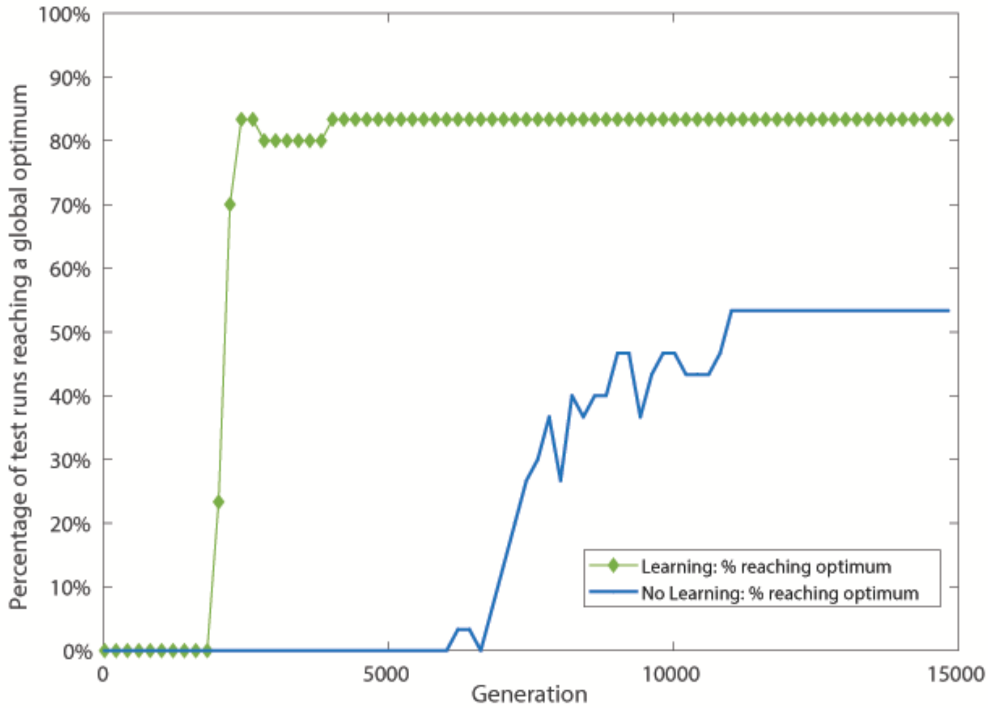


FIGURE 7.6: Comparison of the percentage of runs reaching the global optima at each episode with and without learning for the Modular Constraints (MC) problem. The mean percentage of phenotypes fitness reaching the global optima over the 30 runs is plotted at every 200^{th} generation. The green line with diamond markers shows performance with learning whereas the blue line shows performance without learning. The magnitude of mutation is set to achieve 80% of learning runs reaching a globally optimal fitness with learning.

This result suggests that there is a small differential between the ideal combination of innate behaviours being found with and without learning, and the addition of learning

[3] Assuming evolution alone can hill-climb to a local optima, then the chance of finding one of the two global optima (where there are 10 blocks) is approximately $\frac{2}{2^{10}}$ but is not exactly that figure as basin widths get wider nearer to the global optima.

does not ensure that a global optimum will be always found; this being dependent on mutation rate.

Figure 7.7 shows that with a low enough magnitude for each mutation, the global optimal fitness can be reliably found without learning. This is consistent with the no plasticity result seen in the ESDP model where a small imbalance in the number of correct correlations in random vectors biases selection towards finding a global optimum (see analysis in Section 5.5.3).

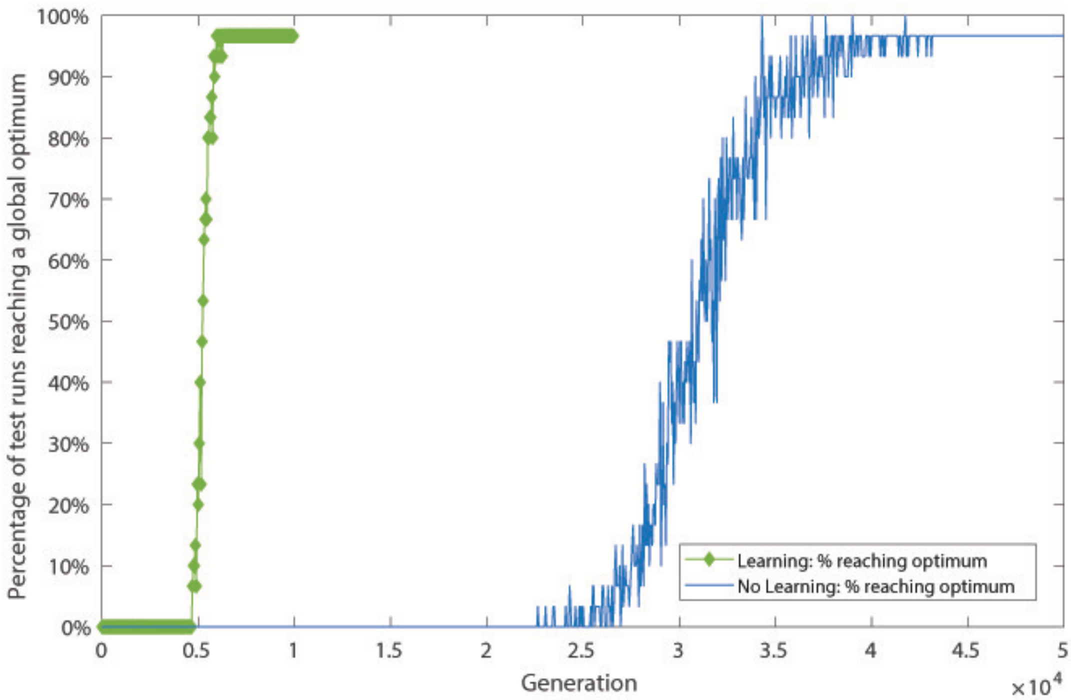


FIGURE 7.7: Percentage of runs reaching the global optima at each episode with and without learning for the Modular Constraints (MC) problem using the highest mutation rate that achieves a 95% or higher of runs reaching the global optima for both learning ($d = 8 \times 10^{-6}$) and no learning ($d = 5 \times 10^{-6}$). The mean percentage of phenotypes fitness reaching the global optima over the 30 runs is plotted at every 200^{th} generation. The green line with diamond markers shows performance with learning whereas the blue line shows performance without learning.

When using the mutation magnitude is tuned to reach the global optimum for 95% or more of the runs, which entails a different rate for learning and no-learning, the speed at which learning reaches the global optimum fitness is approximately seven times faster than without learning. This demonstrates that when considering the time to reach the global optimum, learning significantly speeds the evolution of innate behaviours.

7.4.4 Phenotypic Plasticity is Lost, Suggesting Genetic Assimilation

For this experiment, we test the performance of the simulation with learning to understand if the trial-and-error selection of the environment is being genetically assimilated. This is tested by evolving the genotype with learning and then - once the optimally fit phenotype is produced (after 4000 generations) - switching learning off. As shown by the blue crosses in Figure 7.8, in this case, an optimally fit phenotype is produced at every generation despite the different random start positions at each generation due to the absence of learning. This would suggest that, where learning is deployed, even though the testing of each environment still occurs, these steps become redundant over evolutionary time.

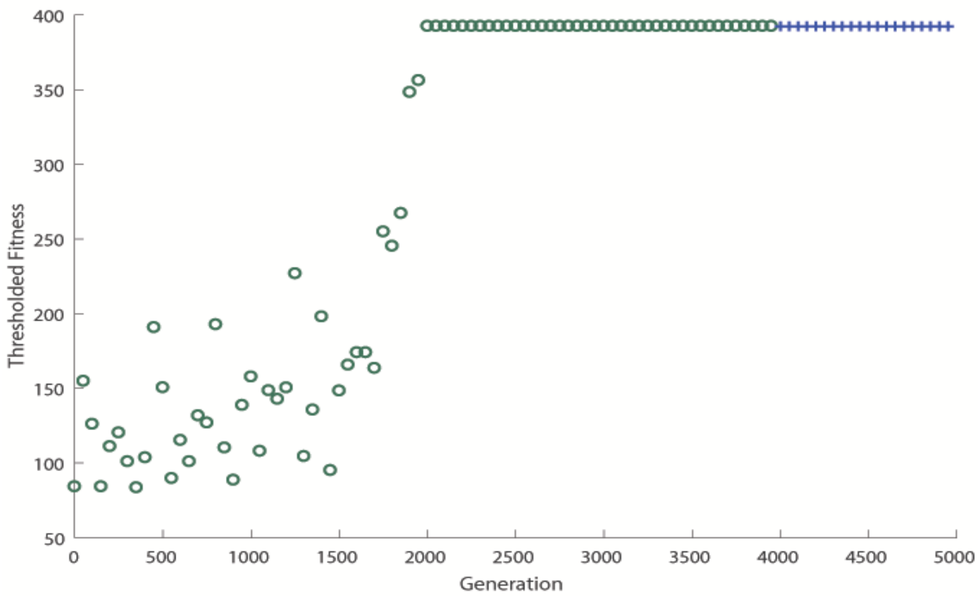


FIGURE 7.8: In this run of the simulation, the first four thousand generations show the action of learning and evolution together: the green circles show the mean phenotype's thresholded fitness (sampled every 50 generations). After 4000 episodes, shown with blue crosses, learning is switched off. The phenotype retains an optimal fitness suggesting that fitness is due to the evolved correlations between innate behaviours.

The loss of plasticity associated with genetic assimilation is further demonstrated by sampling the phenotypes that are produced using the evolved correlations (as encoded by the B matrix) after every thousand generations of evolution and learning. Figure 7.9 shows that at the start of evolution, each environment produces a unique behavioural phenotype. As evolution and learning progress, the range of unique phenotypes being produced significantly decreases, to the point where there at the end of the evolution there are only two unique behavioural phenotypes expressed with, in the majority of runs, globally optimal fitness. Therefore, the degree of plasticity for both learning and innate behavioural expression is almost completely lost. Whether the reduction of plastically expressed phenotypes means that learning is genetically assimilated or genetically accommodated is discussed further in Section 7.5.1.

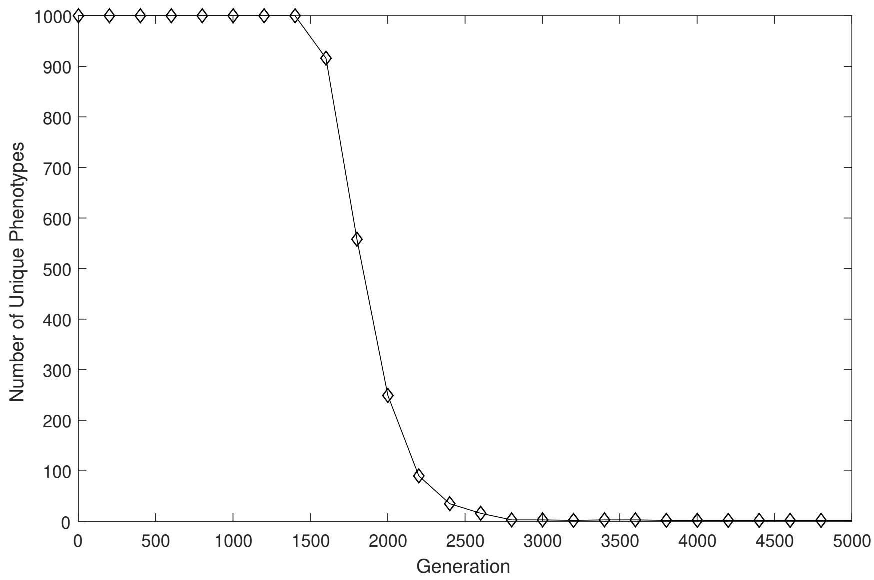


FIGURE 7.9: The number of unique behavioural phenotypes is tracked every 1000 generations for evolution with learning by sampling the matrix of genetic correlations (*B* matrix) and generating 1000 phenotypes from random environmental inputs. To measure the effect of evolution, the sample phenotypes are generated using innate behaviour only. A phenotype is judged to be not unique if the sign of its trait values is the same as the sign of the trait values of another phenotype in the same sample. As evolution progresses, the range of phenotypes generated narrows from each sample phenotype being unique to only two distinct (and globally optimal) phenotypes being produced.

7.4.5 Plasticity is Completely Assimilated in the Absence of Structural Regularities

It is interesting to consider how these results so far reflect on Mery & Burns [77] hypothesis that heterogeneous environments will favour learning over innate behaviours because the behavioural flexibility that learning provides is more adaptive when the environment is uncertain. In this light, one might consider the complete genetic assimilation of learning seen at the end of evolution in the previous experiments with learning a weakness, given that the fitness landscape remains multi-peaked and therefore effectively varying with every generation. However, linkages between innate behaviours have evolved that allow the expression of an optimal phenotype in all environments; the structural regularities in the fitness landscape are mirrored in the genotype. Consequently, additional flexibility through learning is not required. However, the plasticity provided by learning may be advantageous where there are limited structural regularities in the environment.

To test this we use a fitness landscape defined using a constraints matrix *C* matrix, but borrowed from Kounios et al [58], the values of this constraints matrix are set using random values in the range ± 1 except for the self-connections where the value is set to one, so that:

$$C \in \mathbb{R}^{n \times n}$$

$$c_{ij} \sim U([-1,1]), \forall i \neq j,$$

and

$$c_{ij} = 1, \forall i = j$$

In this way, the fitness landscape becomes extremely rugged and lacks the consistent structural regularities of the MC and CC problems.

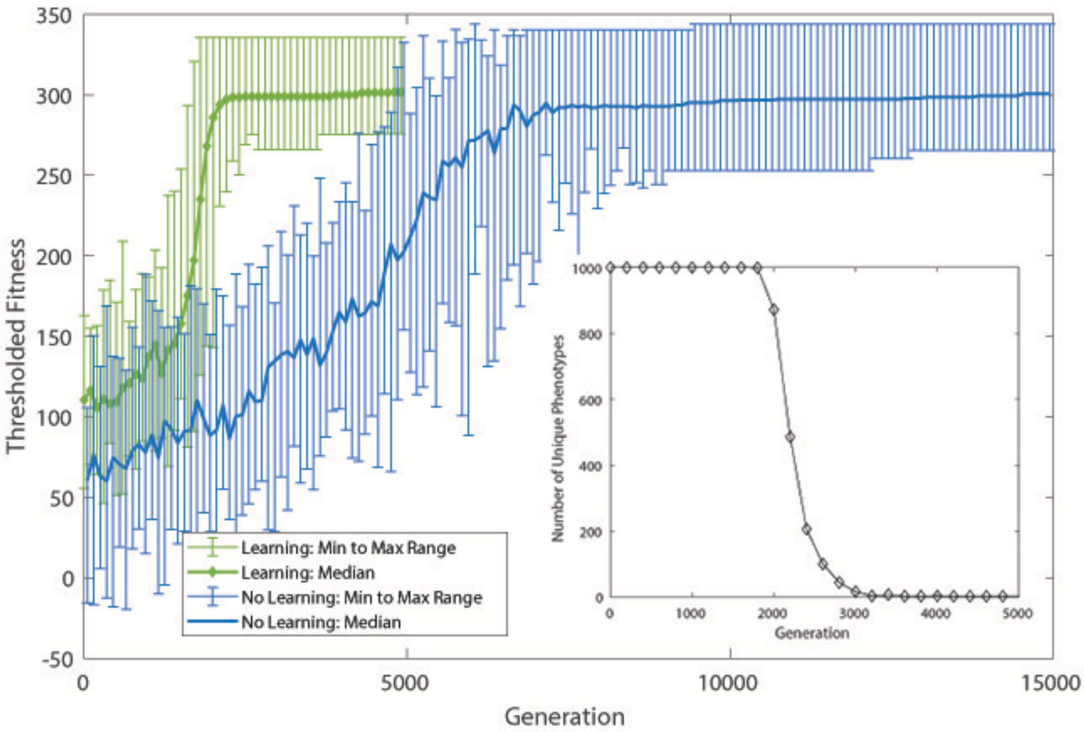


FIGURE 7.10: **Main Figure)** Comparison of fitness with and without learning for the Random Constraints problem. The error bars show the minimum to maximum range of fitnesses achieved at that generation over 30 runs using the standard parameters. The darker line shows the median value over the same 30 runs. The data is plotted at every 100 generations, with the second plot offset by 50 generations. With learning, shown in green with diamond markers and without learning, shown in blue (no markers). **Insert)** As per Figure 7.9, the number of unique innate behavioural phenotypes generated over evolutionary time (with learning).

Figure 7.10 shows that even in the absence of consistent structural regularities, learning speeds the rate of evolution. However, as shown in the inset of Figure 7.10, in common with the MC and CS problems, the number of behavioural phenotypes generated reduce to two over evolutionary time. Given the stochastic nature of the constraints - which are often conflicting - determining closeness to a globally optimal phenotype is problematic and so it is not possible to judge to what extent the innate behavioural phenotypes at the end of evolution are the best average behavioural phenotypes for all environmental inputs. Interestingly, the no learning case appears to achieve a higher fitness than the learning case. This is likely to be because the weaker fitness signal from no-learning provides more time for evolution to resolve conflicting constraints.

Given that, as discussed in Section 2.3.5 in Chapter 2, one might not expect complete assimilation of the learning, the fact that at the end of evolution only two phenotypes are produced in problems with structural regularities and those random constraints may be seen as a model weakness. This result illustrates that there is no mechanism to prevent the continued canalisation of the genotype because a mutation is only being judged on its immediate fitness effects rather than whether this will be detrimental to long-term fitness due to loss of plasticity provided by learning.

7.4.6 Evolution of Innate Behaviours Refines Learning

It is clear from the previous results that learning has a positive impact on the evolution of innate behaviours. But it is also worth considering whether, and how, the evolution of innate behaviours is influencing learning. To that end, we track to what extent the learning is selecting the environment based on the match of the within-block or between-block correlations to the optimal phenotype and how this changes over evolutionary time.

Figure 7.11 shows which environment has been selected based on one of two orderings of the set of environments. The light green line with squares shows the ordering of the selected environment based on the number of correct within block correlations, where the environment with the least correct in-block correlations would be assigned the rank of one and the environment with the most correct within block correlations would be assigned the rank of twenty.

The rank of the environment chosen by learning is calculated by a function that takes the correlation scores of a set of m test environment vectors $g(\boldsymbol{\epsilon}^m)$ and indexes them in ascending order by the number of correct correlations, where the same index is assigned where correlation scores are equal. The result is the index of the environment selected by learning. A correct correlation is defined as when the behavioural environmental elements value e_i and e_j have the same sign (positive and negative) of correlation as the element of the MC constraints matrix C_{ij}. So, for any given environmental vector on n elements, the function g calculates the number of correct correlations, as below:

$$g(\boldsymbol{\epsilon}) = \sum_{i}^{n}\sum_{j}^{n} \begin{cases} 1, & \text{if } e_i e_j C_{ij} M_{ij} > 0; \\ 0, & \text{if } e_i e_j C_{ij} M_{ij} < 0. \\ 0, & \text{otherwise} \end{cases} \tag{7.5}$$

Where, if assessing in-block correlations, the matrix M ($M^{n \times n}$) contains the value one for all within-block correlations and zero otherwise or, when assessing between-block correlations, each element contains the value one for between-block correlations and zero otherwise.

The dark green line shows the same ranking but, this time, based on the number of correct between block correlations.

Here we can see that initially environments with high numbers of correct within block correlations are favoured although this effect quickly reduces to favouring environments with higher numbers of correct between block correlations. So, once the innate behaviours are reaching locally optimal combinations of innate behaviours (at around generation 2000), the effect of the trial-and-error learning is to switch from choosing environments that have high numbers of correlations within a block to choosing environments that have a high number of correct correlations between blocks. Without external intervention, the criteria for successful learning has switched.

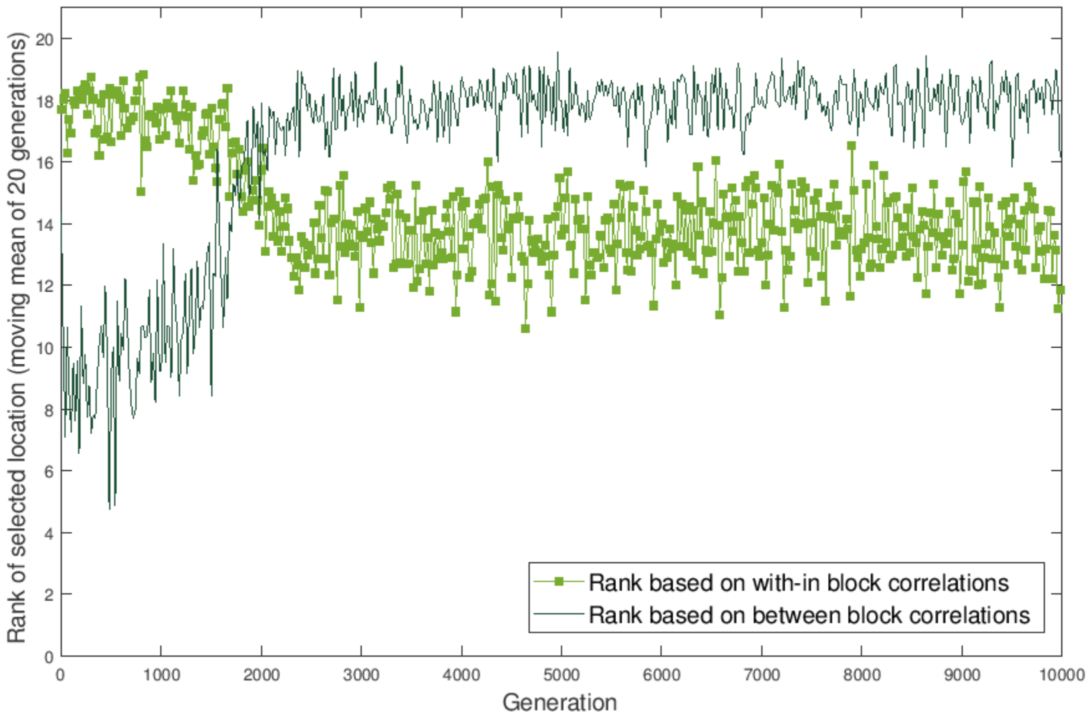


FIGURE 7.11: Index of environment chosen by learning at each generation, where each test environment is sorted in ascending order based on the number of correct correlations within a block (light green line with square markers) and sorted again based on the number of correct correlations between blocks (dark green line). Therefore, the environment with the lowest number of correct correlations has the lowest score, whereas the highest number of correct correlations is ranked highest. The results are smoothed by taking the rolling average across 20 generations. Sample data is plotted every 20 generations.

7.5 Analysis and Discussion

7.5.1 Has Learning Become Genetically Assimilated?

Whilst the Environment Selection model does not exhibit a Baldwinian Optimizing effect - learning is not helping evolution to get to places on the fitness landscape that could not be discovered without learning - this model does raise interesting questions.

In Section 7.4.4, it was suggested that the learning had become genetically assimilated because the range of phenotypes being expressed as evolution progressed was reducing and by the end of evolutionary episodes, either one of two globally optimal phenotypes were consistently produced. Therefore, plastic expression of alternative phenotypes has been limited to two phenotypes. In addition, as Figure 7.12 shows, optimal phenotypes are being discovered by learning before they are being expressed innately in the absence of learning, suggesting that what is being discovered by learning is becoming genetically specified. These are two indicators that learning is being assimilated. But, there are two attributes of this model that brings into question the nature of the assimilation that is taking place.

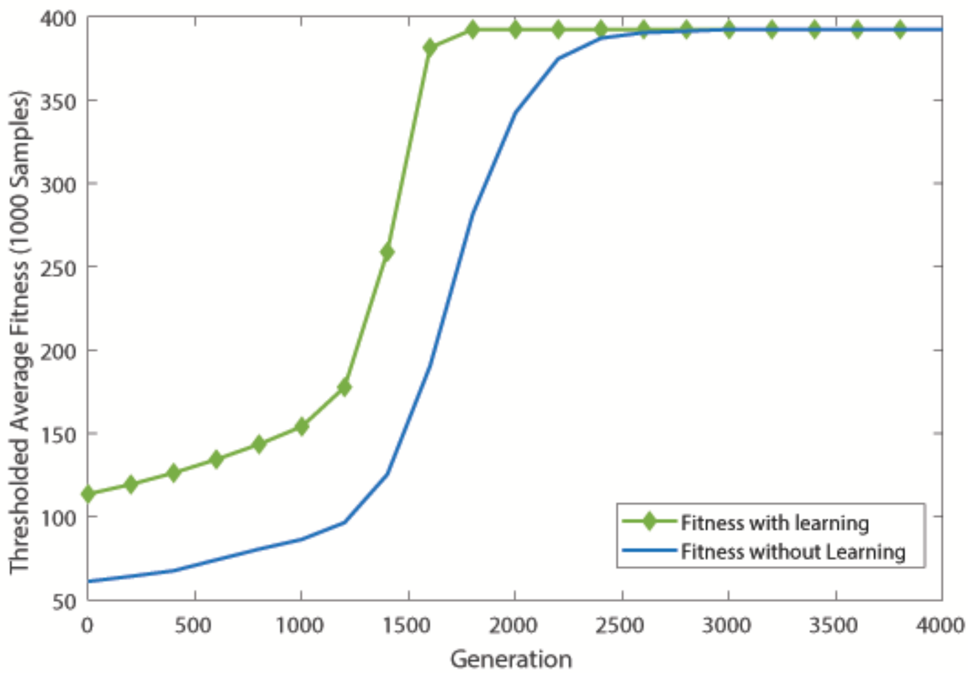


FIGURE 7.12: The average fitness of 1000 phenotypes generated using the correlations matrix B sampled every thousand generations for the MC problem. The phenotype's fitness without learning is shown in blue and the phenotype's fitness with learning is shown in green with diamond markers.

Firstly, in this model, and unlike prior models, learning and evolution are not operating in the same parametric space; genetic evolution is operating in 'correlation space', the phenotype is expressed in 'trait' space and the learning is operating indirectly through the environment. In prior models of learning and evolution, the genetic assimilation is judged to have occurred when the genetic configuration becomes the same as, or near to[4], the configuration of the learnt phenotype. Because learning and evolution are operating in different spaces, these measures cannot apply to this Environment Selection model; the genotype cannot be the same as the phenotype.

Secondly, in this model, there is no static learnt target to which the genetic configuration can assimilate: throughout all evolutionary episodes, the learning is largely stochastic with the environment selected by learning changing with each generation. At the start of evolution, the learnt phenotypes produced are largely random. Consequently, even if the genotype was represented in the same space as the phenotype, there are no fixed learnt traits to which the genome can assimilate. Therefore, one cannot judge whether assimilation has taken place by comparing genetic and phenotypic configurations. One might take the view that the genetic configuration is assimilating to the average phenotype expressed over the generations; those averages being encoded by the genetic correlations between innate behaviours.

The two factors discussed above support the view of learning as improving the signal between variation and selection, rather than being a mechanism for the genotype to 'catch-up' with the phenotype. Learning, in this model, is improving the signal to selection by finding environments where the expression of the innate behaviour is fittest. This to some extent echos the biological reality, whilst overlapping, the neural circuitry for innate behaviours is different to that controlling learning [46] suggesting that they are controlled by different genetic loci.

7.5.2 Applicability Beyond Habitat Selection

The model of learning and evolution presented in this chapter is framed in terms of the basic behaviour of habitat selection: switching habitats changes the expression of innate behaviours which, in turn, changes the rate of evolution. But the model is not specific to just an organism learning in which environment it is best to reside. Consider the bee foraging behaviour example discussed in Section 4.2.2, where scouting, resource selection, and other complex behaviours were dictated by combinations of single behaviours such as sonication and flower selection. The overall behaviour in that example is a mix of individual behaviours in a static environment (the hive) and behaviours exhibited in different environments (foraging patches). Putting aside the hive specific behaviours (which may be viewed as constant), if one assumes that once a

[4] In population based models of learning and evolution, full assimilation tends not to occur because selection pressures do not completely remove mutation effects; for example in the Hinton & Nowlan [49] plastic alleles remain present in the population.

number of foraging patches have been visited, one patch is chosen for resource collection, then the fitness contribution of that bee[5] will be determined by the learnt foraging patch selection. There is a difference in the two scenarios as one can easily imagine that the next foraging trip selects from a different sample of foraging patches and therefore the environmental variability is more fine-grained[6], but this is not expected to qualitatively change the result[7]. The Environment Selection model is still very much a simplification of the actual behaviours exhibited by bees when foraging - both in terms of its gross assumptions (e.g. visiting a set number of environments before choosing one), the treatment of bee's fitness contribution to the population fitness, and in terms of the simplistic modelling of environmental triggers to behavioural responses. Nevertheless, one can see the model applying at the conceptual level to any learning that temporarily alters the experienced environment in which fitness is judged.

Is selecting between environments fundamentally different learning that specifically responds to a single environmental challenge such as the opening of a nutritious nut example used by Heyes, Charter & Dwyer [48]? One could argue that learning to open a nut is the same as moving to an environment where nuts don't have shells. The learning has changed the environment and therefore the selective pressures on the genotype. However, as discussed in Section 2.2.4, moving environments is likely to change the selection pressures on a whole range of potentially unconnected behaviours. For example, moving to a more shaded environment may change the temperature, amount of light and resource availability and therefore could alter selection pressures on many morphological and behavioural traits, e.g. the amount of fur, visual abilities, foraging preferences and predator avoidance. Whereas opening a nut is likely to create selection pressures on a narrower set of behavioural and physical traits.

7.5.3 Direct versus Indirect Learning

The results detailed in the previous section show that at the chosen mutation rate, the indirect action of learning is markedly expediting the progress of evolution (Figure 7.3). The more environments explored by learning the greater the expediting effect although the incremental benefit of exploration rapidly reduces (Figure 7.5). However, there is no evidence of an optimizing effect: using a low mutation rate, evolution is consistently reaching one of the global optima both with and without learning (Figure 7.7). This suggests the indirect nature of the learning means that it has a relatively weak effect on long-term evolutionary outcomes: choosing from one of a random set of environments will provide a stronger fitness signal than from a random environment but a weaker

[5]The analogy is complicated by the fact that bees are social organisms and their activities contribute to the fitness of the collective.

[6]Fine-grained environmental variability is commonly defined as when the environmental changes within an individual organisms lifetime [110].

[7]It is anticipated that the fitness would be the average fitness of the best patches visited and this would likely dilute the impact of selecting the best of multiple environments.

signal than a locally optimal environment, i.e. one that mirrors a locally optimal phenotype. Since the choice of which random environments to explore is independent of the innate behaviours - which is not the case with the Correlated Behaviours model - the chances of selecting a locally optimal environment when visiting twenty environments is vanishingly small and in the order[8] of ($20 \times \frac{2^{10}}{2^{60}} = \frac{20}{2^{50}}$).

Although the algorithms are not directly comparable - it is worth considering how the weak signal improvement of the Environmental Selection model compares to that of the Correlated Behaviours model. Learning in the Environmental Selection model changes the start conditions (i.e. the environmental input) for the expression of innate behaviours, whereas in the Correlated Behaviours learning changes the innate behaviour's direction (encoded as the sign of the trait). For both models at the end of learning, the fitness of the mutation to the correlation strength is assessed in the context of the learnt phenotype being closer to a local optimum. In the Correlated Behaviours model, at the start of evolution, the innate behavioural expression is at a locally optimal input - learning changes the sign of behavioural traits so that they are consistent within a block. In the Environmental Selection model, learning would on average need to visit 2^{50} environments for there to be an equivalent, locally optimal, start condition for the innate behavioural expression. However, as shown in Figure 7.5, in practice, the incremental effect of increasing the number of environments explored diminishes rapidly, in practise the number of environments needed for the Environmental Selection model to generate the same fitness signal as the Correlated Behaviours model is likely to be significantly less than 2^{50}.

This suggests it is the constraint of the search space through the genetic linkages in the Correlated Behaviour model that adds significantly to the expediting effect rather than the assessment of the mutation in a better quality context as provided by a local optimum. However, since improvement of the quality of the fitness signal is the only method that learning can improve evolution, and the magnitude and rate of mutation dictates how quickly a genotype can change, it would seem that, for the Correlated Behaviour model, most of the work in the rapid discovery of global optima is due to the action of learning itself.

7.6 Conclusion

The Environment Selection correlation-based model presented in this chapter shows that learning which alters the environment experienced by the phenotype both speeds evolution and improves the reliability at which a global optimum fitnesses can be found for a given mutation magnitude. This result holds for both the MC and CC problem structures and is therefore not problem-specific. Learning can also speed the rate of

[8]The actual probability is $1 - (\frac{(2^{60}-2^{10})}{2^{60}})^{20}$ - it is hard to gauge the magnitude from this equation.

evolution for random constraints that generate a highly rugged, unstructured fitness environment with evolution canalising the range of behavioural phenotypes produced.

In addition, this model significantly relaxes the conditions under which genetic assimilation of learnt behaviours takes place, as discussed in Chapter 3. Firstly, the model does not require a cost of learning for the learning to be assimilated. Secondly, the parametric space of movement of the phenotype (trait space) is different to that of the genotype (correlation space) and the environment ('environment' space) therefore there is little scope of meaningful neighbourhood correlation. Thirdly, there is no consistent learnt outcome to which the genome can assimilate. The results clearly show that through the evolution of genetic correlations, there is a loss of plasticity for both the learnt and innate behaviours, with the genetic constraints enabling the expression of globally optimal behavioural phenotypes.

Chapter 8

Conclusions

As summarised in Figure 2.4 in Chapter 2, the routes by which learning and evolution interact are many and various, with the potential for learning to impact selection both by directly changing behavioural trait values and through altering the environment that an individual or population experiences. Further complexity is added to this relationship by the potential for learning to influence inheritance and for learning to persist across generations through social effects. The review of prior literature and theoretical models in Chapter 2 identified Waddington's [119] concept of genetic assimilation via canalisation of the genotype due to pleiotropic and epistatic effects as a candidate mechanism underpinning the Baldwin Effect. This canalisation model of the Baldwin Effect is appropriate to the relationship between learning and innate behaviours, due to the potential for learning to utilise genetically linked sets of innate behaviours. This interaction has not been previously modelled.

In Chapter 3, we explored the necessary conditions for learning to be assimilated to become innate behaviour. Exploration of the dynamics of prior simulations, where both learning and evolution operate in 'trait space', suggests the need for a cost of learning, a similar fitness effect between the movement in genotype space (mutations) and movement in phenotype space (learning steps) as well as stability in the outcome of the learning. Whether these conditions are necessary under all circumstances is nuanced with exceptions for cases such as drift as well as dependencies on the frequency of environmental change. Part of the ambition of this thesis was to understand if these conditions are necessary when evolution operates in 'correlation space', i.e. genetic linkage between innate behaviours. Viewing learning in terms of the fitness signal between variation and selection helps the understanding of when and why these conditions apply.

Chapter 4 introduced prior model problems used to assess the performance of the simulations presented in this thesis. The case was then made for why these model problems have relevance for a model of behaviour where good combinations of behaviours have locally optimal fitness and ideal combinations of behavioural expression have globally optimal fitness.

Whilst modelling development plasticity rather than learning, the Environmentally Sensitive Development Plasticity model presented in Chapter 5 demonstrated that plasticity can be assimilated where that plasticity guides the canalisation of the genotype. Importantly, in this model both the adaptive plasticity and the evolution operated in correlation space; the plasticity was incremental and so avoided the scenario where learning produces the same phenotype before and after a mutation and therefore has the potential to 'hide' the fitness signal from evolution. However, whilst the plasticity case did not get stuck at a local optimum and therefore reached the global optima quickly, with certain model parameters, the no learning case also reliably found a global optimum, albeit more slowly.

Chapter 6 considered a Correlated Behaviours model of learning and evolution that allowed a generative learning mechanism to change the expression of innate behaviours based on the genetic linkage between those behaviours. The evolved linkages allowed the learning to scale-up, first altering single behavioural traits and then sets of closely correlated behavioural traits. Importantly, in this model evolution operates in correlation space whereas learning uses correlation space to search in trait space. In common with the ESDP model, the plastic expression of the phenotype was assimilated but in this case, the rate of that assimilation increased as learning scaled up to alter multiple connected behavioural traits. The no-learning case did not reliably discover the globally fit phenotypes and therefore a Baldwin Optimizing Effect was observed.

Another variation of the model using a similar structure to the ESDP and Correlated Behaviours models was devised in Chapter 7; the Environment Selection model enabled learning to choose in which environment an individual should reside. This indirect mechanism of learning changes selection pressures on the genotype by altering the environment that the individual experiences. Interestingly, at the start of evolution, learning chose environments that favoured correct within block constraints and then, as evolution progressed, favoured environments with more correct between-block connections. Whilst it is not claimed that this is a self-rescaling of learning, it does show that a genetic correlation between innate behaviours can spontaneously alter the choices that learning makes.

Together, the Correlated Behaviours and Environment Selection models have demonstrated, through a variety of causal routes and parametric spaces, a two-way feedback between learning and evolution where learning guides the canalisation of the genotype and this genotype channels what is learnt. Further, correlation models of learning and the multiple ways in which learning can influence selection, suggests that the conditions under which genetic assimilation can occur are broader than the analysis of traditional trait focused models would suggest.

In this thesis, evidence for the claims was provided as follows:

1. ***Canalisation of the genotype can be guided by learning (a) directly, when the action of learning alters the expression of innate behaviours or (b) indirectly, when learning alters the experienced environment.*** A direct causal route for learning guiding canalisation to discover globally optimal phenotypes was demonstrated by the Correlated Behaviours model whereas an indirect route to globally fit phenotypes was shown in the Environmental Selection model.

2. ***Direct two-way feedback between learning and evolution can change the movement of the phenotype in the fitness landscape in a way that allows rapid discovery of globally optimal fitness phenotypes that are unlikely to be reliably found by evolution alone.*** The Correlated Behaviours model demonstrated a two-way interaction between learning and evolution that reliably reached global optimal fitnesses, whereas without learning did not.

3. ***The conditions normally associated with the genetic assimilation of learning are relaxed when that assimilation is achieved by the genetic linkage between innate behaviours.*** Both the Correlated Behaviours and Environment Selection model showed that learnt behaviours can be assimilated in the absence of a cost of learning, with a large disparity between movement in genotype space and phenotype space and without a stable learning target to which to assimilate.

8.1 Implications for Phenotype-first Evolution

There is much debate between the Modern Synthesis view of evolution and the emerging Extended Evolutionary Synthesis (EES) as to the importance of the role of the phenotype in dictating evolutionary outcomes [62]. Whilst the arguments in this debate are nuanced and exploring them is a thesis in its own right, it is worth briefly considering if this work has any implications for a phenotype-first view of evolution.

To crudely summarise the debate using extremes, the traditionalists' Modern Synthesis view would be that whilst morphological and phenotypic plasticity have a role in shaping the phenotype - gene frequencies in a population dictate that plasticity and consequently all evolutionary outcomes are just a product of the genes [127]. The EES counter to this argument is that, if you take into account the multitude of non-genetic influences on the phenotype - including behaviour - an individual's genome is just one of many formative influences and therefore the phenotype is under-determined by the genotype [61]. The debate to some extent becomes a battle of population-level versus individual-level explanations of evolutionary outcomes.

Whilst one might argue that innate behaviours, whilst environmentally sensitive, are genetically specified, it is more difficult to make the same case for learning. Indeed, I would argue that learning is a linchpin for the phenotype-first view of evolution because unlike developmental plasticity and innate behaviours, learning carries a *'state'* from its history of complex interactions with an ever-changing environment. It is this state that makes the mapping between genotype and 'learnt phenotype' so complex that allele frequencies in the population quickly become an unsatisfactory level of explanation for a given phenotype's fitness.

Do the results from the experiments presented in this thesis tell us anything new about phenotype-first evolution? The results show that plasticity in general and learning in particular significantly accelerates the progress of evolution. That is a common result. However, the Correlated Behaviours model also shows a Baldwin Optimizing effect: the two-way feedback loop between learning and evolution is finding phenotypes in the fitness landscape that would not be reliably found without learning. Does this in itself make learning special? There are prior models of evolution that demonstrate an optimizing effect where local optima are escaped through canalisation of the genotype without any form of plasticity. For example, Kounios et al.'s [58] model of evolvability demonstrates evolution reaching a globally optimal configuration in a rugged fitness landscape using a phenotypic development mechanism that does not include developmental plasticity. In Kounios et al.'s model and those presented in this thesis, escape from a local optimum is driven by the evolving correlations in the genome: not learning. In addition, there are other dynamics that find the global optimum in a multi-peaked fitness landscape - for example, the ratcheted drift in Borenstein et al.'s [12] model discussed in Chapter 3.

However, the work of this thesis does help cement the view of learning as a vital actor in evolution in two ways. Firstly, the bi-directional interaction between learning and innate behaviours as demonstrated in the Correlated Behaviours and Environment Selection models shows that the effect of learning may be more complex than often assumed in prior models: learning doesn't find ideal phenotypic configurations, as it is often modelled in prior literature - learning changes the adaptive fitness signal to evolution both directly and via the environment. The evolved behaviours then channel what can be learnt creating a complex dynamic that cannot be predicted without considering the entire history of interactions between the genetic disposition and environmental factors. Secondly, the models presented in this thesis demonstrate that learning can be assimilated without the need for an explicit cost of learning and where the learning changes frequently - conditions usually thought necessary for assimilation. This suggests that learning's effect on evolution can be long lasting in a broad set of circumstances.

8.2 Future Directions

The work of this thesis suggests learning, innate behaviours and physical traits are all important to shaping evolutionary outcomes. Whilst all three elements have been included in at least one model, no model has been constructed of morphological traits interacting with learning (potentially in concert with developmental plasticity). This is likely to be an important aspect of how learning and evolution interact; both in terms of West-Eberhard's concept of physical trait-behavioural complexes (discussed in Section 2.7.4) and Heyes et al.'s [48] suggestion that it is peripheral mechanisms of learning including physical sensory systems that have evolved via the Baldwin Effect. A potential future direction is to implement models of these effects with a similar structure to the ones presented in this thesis.

In addition, there are several areas where there has not been sufficient time to go beyond a preliminary investigation and are therefore candidates for future work. These are:

- **Neighbourhood correlation in complex G-P maps.** In Mayley's original paper, the concept of neighbourhood correlation and its requirement for genetic assimilation is presented in terms of the correlation of the size of movement in genotype space being commensurate with the size of movement in phenotype space. Further exploration of the incomplete assimilation in Mayley's model shows that assimilation is dependent on the fitness effects of individual learning steps versus the fitness effects of single or multiple mutations - rather than the total size of movement in phenotype space. However, as Mayley also identifies, there is potential for the assimilation to be frustrated when the genotype is close to one adaptive peak and the phenotype is close to a different peak. The random nature of the NK landscape in the original model does not aid analysis of this issue as it is difficult to judge the basin of attraction for any adaptive peak. Preliminary investigations using the MC landscape suggest that a cost of learning can pull the genotype across a basin of attraction, but this is yet to be rigorously tested. In addition, developmental processes causing complex genotype-phenotype maps may constrain which phenotypes are accessible through evolution potentially also inhibiting genetic assimilation. This would be an interesting area for analysis as it could potentially recast the neighbourhood correlation challenge in terms of accessibility rather than the size of movement.
- **The impact of social transmission on a canalisation model of assimilation.** All the models in this thesis assume that there is no transmission of learning between generations. Whilst this broad assumption helps isolate and exemplify the action of genetic assimilation of the learning, it largely ignores the biological reality that learning is transmitted between generations by mechanisms such as parental

teaching. It is unclear how social transmission would inhibit or potentially accelerate genetic accommodation or assimilation of the learning, in the context of models where learning directly or indirectly coordinates innate behaviours.

- **Learning altering the environment or inheritance.** As identified in Panel E of Figure 2.4, there are many causal paths for the interaction of learning and evolution that have not been considered in this thesis. Examples include perturbational niche construction where the constructed niche may stabilize environmental inputs across multiple generations and sexual selection, which is often learnt, is likely to have a cost of learning and can also lead to run-away evolutionary effects [118]. Testing these routes using a model where learning facilitates correlation between innate behaviours is likely to be a fruitful area for investigation.

- **Correlated innate behaviours and genetic connection structures.** The models presented in this thesis are reliant on the correlations evolved generalising the regularities in the fitness landscape, where there are sets of strongly correlated behaviours linked together by weaker connections. It is conceivable that innate behaviours may have more complex interrelations than those simulated in this thesis, for example, sets of strongly correlated behaviours might be overlapping. An interesting broad research question raised by this work is to what extent there is a need for more complex genetic connection structures to represent potentially more complex connections between innate behaviours and does this limit how complex an innate behaviour can become.

It is clear from the above, that there is some way to go before there is a comprehensive understanding of how learning and evolution interact. In addition, learning is one example of a biological *exploratory process* [62] that exhibit open-ended behaviour which alters the fitness of the phenotype. Other examples include the immune system, which matures over time [55, 131] (learns), and how a plant's roots explore the soil to find moisture [42]. Given the importance of these exploratory processes to the development of a phenotype-first framework for evolution, this thesis provides a small contribution to improving the collective understanding of this domain.

Appendix A

Models of Canalisation Through Epistatic Effects

The modelling work in this thesis draws some inspiration and has structures in common with three models that examine the effects of canalisation and these are briefly surveyed in this Appendix.

The first model of note was that of Siegal & Bergman [105]. Their work suggested that gene transcription factors combined with a development process will cause robustness against mutation and environmental perturbations even when there is no stabilising selection towards a fitness optimum. Their simulation represented a network of genes as a matrix where the gene expression at a given time is defined by a state vector (S). A phenotype was developed by iterating the state vector until an equilibrium was reached (i.e. until the values in the state vector did not change significantly with each iteration). Unstable individuals, ones that did not become stable after a set number of iterations, were assigned a zero fitness, whereas stable individuals were assigned a fitness based on the Hamming distance to a randomly chosen target phenotype. The authors tested different selection strengths and found that while increasing selection strength did speed the rate at which individuals reached a stable state while increasing resistance to mutation, a zero selection strength also leads to increasingly robust and stable phenotypes over evolutionary time. They also found that the more genetic connections that are allowed to evolve, the more the phenotype is robust against mutation. The authors suggest that this helps explain why species are observed to have abundant genetic variation, experience a range of environmental conditions but display limited phenotypic variation (i.e. phenotypic canalisation [63]).

Building on the Siegal & Bergman [105] model, Masel [72] tested whether selection against an optimum phenotype is required for genetic assimilation. The inspiration for this being to test whether assimilation was due to canalisation of the genotype. In this version of the experiment, noise in the development function was used to simulate environmental perturbations (in this case, heat shock). The evolutionary process first selected for a phenotype that was stable at the end of development, followed by stabilising selection towards the target of one developmentally stable phenotype (this canalised the population around the stable phenotype). The next step was to search for a *'phenocopy'*; a phenocopy is where a mutation produces the same phenotypic effect as an environmental perturbation. Once a phenocopy was found; two tests were performed: the first was selection for developmental equilibrium only (i.e. canalisation) and the second was for selection against the target phenotype. Assimilation was assessed by testing whether the correlations matrix produced the same phenotype under high and low noise conditions (i.e. the phenotype was being expressed independently of the environmental input). The key result was that genetic assimilation took place even when only selecting for developmental stability but was more frequently observed when selecting for the target phenotype.

Later a model by Espinosa-Soto et al. [33] examined whether phenotypic plasticity facilitates adaptive evolution in gene regulatory circuits. In common with the models developed by Andreas Wagner and others (cited by Espinosa-Soto), the experimental set-up utilised a gene regulatory network specified by a matrix that represented an individual's genotype. The individual's phenotype was determined by iterating an initial state gene activity pattern (S0) through a thresholded equation of motion until a stable state is achieved, with unstable phenotypes discarded. In their analysis, two types of changes were used: genetic change through the mutation to the elements of the gene regulatory circuit (A matrix) and plasticity through the perturbation of the initial state (S0). The authors conducted a comprehensive analysis of the distribution and properties of the phenotypes produced when there is a perturbation of the initial state (plasticity) and when changes to the gene regulatory circuit were made (genetic mutation). The main result showed that new, adaptive phenotypes resulting from a non-genetic perturbation could be stabilised through selection to become the native phenotype. Therefore, the authors concluded that plasticity is likely to have a strong role in the evolution of beneficial gene activity patterns.

These simulations suggest that canalisation plays an important role in developmental plasticity and therefore is an important consideration for any model of learning and evolution that has a network component model. We apply an equivalent concept of canalisation to behaviours.

Appendix B

Calculation of Signal Strength in Random Vectors

As discussed in Section 5.5.3 of Chapter 5, the signal strength provided by the asymmetry between positively correlated traits being in the majority more often than being in the minority can be calculated, as described below.

For a random vector of n loci that can take either a positive (+1) or negative value (-1), the signal strength Q is calculated as:

$$Q = \sum_{z=0}^{n} X_z Y_z$$

where z is the number of correct loci in comparison to the target vector.

X_z is the expected proportion of vectors that have z correct values. X_z is calculated by dividing the binomial coefficient of z and n by the total number of possible vectors (2^n) to get the proportion of vectors with that number of correct values. Therefore,

$$X_z = \frac{n!}{2^n z!(n-z)!}$$

Where 'correct value' is defined as where a locus matches that in the target vector (e.g. all ones) and excluding the target vectors complement (e.g. all negative ones). The complement case (where there are n-z correct loci) which provides the same signal is accounted for because in this equation z' gives the same result as z where $z' = n - z$. For example, where the vector has five loci, the proportion of vectors with one correct value ($\frac{5}{32}$) is the same as the proportion of vectors with four correct values ($\frac{5}{32}$).

Y_z is the expected probability of the correct signal for vectors with z correct values. The correct signal is one where a change to the connection strength between s_i and s_j is in the same direction of the constraints matrix C (i.e. negative if s_i and s_j are negatively correlated and positive if s_i and s_j are positively correlated). This correct signal occurs when s_i and s_j are of the same sign and agree with the majority of signs within the vector (either positive or negative).

This is calculated by finding the probability that a fitness change agrees with the majority of values in the random vector which is the probability of both loci agreeing with the majority minus the probability of both loci disagreeing with the majority (i.e. being in the minority):

$$Y_z = \left(\frac{1}{2n} \left| n - 2z \right| \right)$$

The derivation of this is below for z correct values in a vector of length n.

For the first loci, the number of values in the majority is calculated as: $|z - \frac{n}{2}| + \frac{n}{2}$ and the probability of a locus value being in that majority is $\frac{1}{n}\big(|z - \frac{n}{2}| + \frac{n}{2}\big)$. Conversely, the number of locus values in the minority is calculated as $\frac{n}{2} - |\frac{n}{2} - z|$ with the probability of being in the minority is $\frac{1}{n}\big(\frac{n}{2} - |\frac{n}{2} - z|\big)$.

For the second locus, the probabilities are different since the number of available loci is reduced by one. Hence, the probability of the second locus value agreeing with the majority is: $\frac{1}{n-1}\big(|z - \frac{n}{2}| + \frac{n}{2} - 1\big)$ and agreeing with the minority is $\frac{1}{n-1}\big(\frac{n}{2} - |\frac{n}{2} - z| - 1\big)$.

The signal is the difference in probability of both values being in the majority minus the probability of both being in the minority. Hence:

$$Y_z = \Big(\big(\frac{1}{n}\big)\big(|z - \frac{n}{2}| + \frac{n}{2}\big)\big(\frac{1}{n-1}\big)\big(|z - \frac{n}{2}| + \frac{n}{2} - 1\big) \Big) - \Big(\big(\frac{1}{n}\big)\big(\frac{n}{2} - |\frac{n}{2} - z|\big)\big(\frac{1}{n-1}\big)\big(\frac{n}{2} - |\frac{n}{2} - z| - 1\big) \Big)$$

Since $|a - b| = |b - a|$, and defining $c = |\frac{n}{2} - z|$ and simplify to:

$$Y_z = \frac{1}{n(n-1)} \Big(\big(\frac{n}{2} + c\big)\big(\frac{n}{2} + c - 1\big) - \big(\frac{n}{2} - c\big)\big(\frac{n}{2} - c - 1\big) \Big)$$

So:

$$Y_z = \frac{1}{n(n-1)} \Big(\frac{4cn}{2} - 2c \Big) = \frac{1}{n(n-1)} \Big(2c(n-1) \Big) = \frac{1}{n} 2c$$

Substituting $c = |\frac{n}{2} - z|$ back:

$$Y_z = \frac{1}{n} |n - 2z|$$

References

[1] D. Ackley and M. Littman. Interactions Between Learning and Evolution. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, Artificial Life II (Santa Fe Institute Studies in the Sciences of Complexity Proceedings), pages 487–509. Addison Wesley, artificial edition, 1991.

[2] I. Adrian-Kalchhauser, S. E. Sultan, L. N. Shama, H. Spence-Jones, S. Tiso, C. I. Keller Valsecchi, and F. J. Weissing. Understanding 'Non-genetic' Inheritance: Insights from Molecular-Evolutionary Crosstalk. Trends in Ecology & Evolution, 35(12):1–12, 2020.

[3] L. W. Ancel. A quantitative model of the Simpson-Baldwin Effect. Journal of theoretical biology, 196(1999):197–209, 1999.

[4] L. W. Ancel. Undermining the Baldwin Expediting Effect: Does Phenotypic Plasticity Accelerate Evolution? Popul. Biol, 319:307–319, 2000.

[5] J. R. Auld, A. A. Agrawal, and R. A. Relyea. Re-evaluating the costs and limits of adaptive phenotypic plasticity. Proceedings of the Royal Society B: Biological Sciences, 277(October 2009):503–511, 2010.

[6] J. M. Baldwin. A New Factor in Evolution. The American Naturalist, 30(354):441–451, 1896.

[7] L. R. Baugh and T. Day. Nongenetic inheritance and multigenerational plasticity in the Nematode C. Elegans. eLife, 9:1–13, 2020.

[8] N. Behera and V. Nanjundiah. An investigation into the role of phenotypic plasticity in evolution. Journal of Theoretical Biology, 172(3):225–234, 1995.

[9] N. Behera and V. Nanjundiah. trans gene regulation in adaptive evolution: A genetic algorithm model. Journal of Theoretical Biology, 188(2):153–162, 1997.

[10] N. Behera and V. Nanjundiah. Phenotypic plasticity can potentiate rapid evolutionary change. Journal of Theoretical Biology, 226(February 2004):177–184, 2003.

[11] D. A. Berg, Howard C; Brown. Chemotaxis in E coli analysed by 3D tracking. Nature, 239:500–504, 1972.

[12] E. Borenstein, I. Meilijson, and E. Ruppin. The effect of phenotypic plasticity on evolution in multipeaked fitness landscapes. European Society for Evolutionary Biology, 19:1555–1570, 2006.

[13] C. A. Botero, F. J. Weissing, J. Wright, and D. R. Rubenstein. Evolutionary tipping points in the capacity to adapt to environmental change. Proceedings of the National Academy of Sciences, 112(1):184–189, 2015.

[14] R. L. Brown. Learning, evolvability and exploratory behaviour: Extending the evolutionary reach of learning. Biology and Philosophy, 28(6):933–955, 2013.

[15] I. Carcea and R. C. Froemke. Biological mechanisms for observational learning. Current Opinion in Neurobiology, 54(December 2018):178–185, 2019.

[16] D. R. Chenard KC. The Special Case of Behavioral Plasticity? In D. Pfenning, editor, Phenotypic Plasticity and Evolution, chapter 12, pages 301–326. Taylor & Francis Group, LLC, Abingdon, 1 edition, 2021.

[17] L. M. Chevin and R. Lande. Evolution of environmental cues for phenotypic plasticity. Evolution, 69(10):2767–2775, 2015.

[18] M. H. Christiansen, F. Reali, and N. Chater. The Baldwin Effect Works for Functional, But Not Arbitrary, Features of Language. In Proceedings of the 6th International Conference (EVOLANG6), pages 27–34, 2006.

[19] N. Claidière and D. Sperber. Imitation explains the propagation, not the stability of animal culture. Proceedings of the Royal Society B: Biological Sciences, 277(1681):651–659, 2010.

[20] D. Clarke and D. R. Obert. Predictive modelling of honey bee foraging activity using local weather conditions. Apidologie, (49):386–396, 2018.

[21] P. A. Corning. Evolution 'on purpose': How behaviour has shaped the evolutionary process. Biological Journal of the Linnean Society, 112(2):242–260, 2014.

[22] E. Crispo. The Baldwin Effect and genetic assimilation: Revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. Evolution, pages 2469–2479, 2007.

[23] A. Crombach and P. Hogeweg. Evolution of evolvability in gene regulatory networks. PLoS Computational Biology, 4(7), 2008.

[24] E. Danchin, A. Pocheville, O. Rey, B. Pujol, and S. Blanchet. Epigenetically facilitated mutational assimilation: epigenetics as a hub within the inclusive evolutionary synthesis. Biological Reviews, 94(1):259–282, 2019.

[25] G. K. Davis and M. A. Wund. Developmental Plasticity and Phenotypic Evolution. Encyclopedia of Evolutionary Biology, 1(1):430–440, 2016.

[26] T. J. DeWitt, A. Sih, and D. S. Wilson. Costs and limits of phenotypic plasticity. Trends in Ecology & Evolution, 13(2):77–81, 1998.

[27] R. Diogo. Evolution Driven by Organismal Behavior: A Unifying View of Life, Function, Form, Springer, Cham, 2017.

[28] H. Dopazo, M. B. Gordon, R. Perazzo, and S. Risau-Gusman. A model for the interaction of learning and evolution. Bulletin of Mathematical Biology, 63(1):117–134, 2001.

[29] K. L. Downing. Development and the Baldwin Effect. Artificial Life, 10(1):39–63, 2004.

[30] J. A. Draghi and G. P. Wagner. The evolutionary dynamics of evolvability in a gene network model. Journal of Evolutionary Biology, 22:599–611, 2009.

[31] J. A. Draghi and M. C. Whitlock. Phenotypic Plasticity Facilitates Mutational Variance, Genetic Variance, and Evolvability Along the Major Axis of Environmental Variation. Evolution, pages 2891–2902, 2012.

[32] S. Dridi and L. Lehmann. Environmental complexity favors the evolution of learning. Behavioral Ecology, 27(3):842–850, 2016.

[33] C. Espinosa-Soto, O. C. Martin, and A. Wagner. Phenotypic plasticity can facilitate adaptive evolution in gene regulatory circuits. BMC Evolutionary Biology, 11:5, 2011.

[34] J. F. Fontanari and R. Meir. The Effect of Learning on the Evolution of Asexual Populations. Complex Systems, 4:401–414, 1990.

[35] S. A. Frank. The Design of Natural and Artificial Adaptive Systems. in Adaptation, M.R. Rose and G.V. Launder, eds., pages 451–505, 1996.

[36] S. A. Frank. Natural selection. II. Developmental variability and evolutionary rate. Journal of Evolutionary Biology, 24(11):2310–2320, 2011.

[37] R. M. French and A. Messinger. Genes, phenes and the Baldwin effect: learning and evolution in a simulated population. Artificial life IV. Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living SystemsLife {IV}, pages 277–282, 1994.

[38] C. K. Ghalambor. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Functional Ecology, pages 394–407, 2007.

[39] C. K. Ghalambor, L. M. Angeloni, and S. P. Carroll. Behaviour as phenotypic plasticity. In D. Westneat and C. Fox, editors, Evolutionary Behavioural Ecology, number December, pages 90–107. Oxford University Press, 1 edition, 2010.

[40] C. K. Ghalambor, K. L. Hoke, E. W. Ruell, E. K. Fischer, D. N. Reznick, and K. A. Hughes. Evolution of gene expression in nature. Nature, 2015.

[41] J. Gillespie. Molecular Evolution Over the Mutational Landscape. Evolution, 38(5):1116–1129, 1984.

[42] S. Ginsburg and E. Jablonka. Epigenetic learning in non-neural organisms. Journal of Biosciences, 34(4):633–646, 2009.

[43] D. E. Goldberg and J. H. Holland. Genetic Algorithms and Machine Learning. Machine Learning, 3:95–99, 1988.

[44] J. Gould and P. Marler. Learning by Instinct. Scientific American, 256(1):74–85, 1986.

[45] S. Gould. The Structure of Evolutionary Theory. Belknap Harvard, 2002.

[46] I. C. Grunwald Kadow. State-dependent plasticity of innate behavior in fruit flies. Current Opinion in Neurobiology, 54:60–65, 2019.

[47] B. K. Hall. Baldwin and Beyond: Organic Selection and Genetic Assimilation. In B. H. Weber and D. J. Depew, editors, Evolution and Learning: The Baldwin Effect Reconsidered, pages 141–167. MIT Press, 2003.

[48] C. Heyes, N. Chater, and D. M. Dwyer. Sinking In: The Peripheral Baldwinisation of Human Cognition. Trends in Cognitive Sciences, 24(11):884–899, 2020.

[49] G. E. Hinton and S. J. Nowlan. How Learning Can Guide Evolution. Complex Systems, 1:495–502, 1987.

[50] G. R. Hunt and R. D. Gray. Genetic assimilation of behaviour does not eliminate learning and innovation. Behavioral and Brain Sciences, 30(4):412 to 413, 2007.

[51] E. I. Jones and A. Dornhaus. Predation risk makes bees reject rewarding flowers and reduce foraging activity. Behavioral Ecology and Sociobiology, 65(8):1505–1511, 2011.

[52] N. Kashtan and U. Alon. Spontaneous evolution of modularity and network motifs. Proceedings of the National Academy of Sciences, 102(39):13773–13778, 9 2005.

[53] N. Kashtan, E. Noor, and U. Alon. Varying environments can speed up evolution. Proceedings of the National Academy of Sciences of the United States of America, 104(34):13711–13716, 2007.

[54] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. Journal of Theoretical Biology, 22(3):437–467, 1969.

[55] S. A. Kauffman and S. Levin. Towards a general theory of adaptive walks on rugged landscapes. Journal of Theoretical Biology, 128(1):11–45, 1987.

[56] S. Kaufman. The Origins of Order: Self organisation and selection in evolution. Oxford University Press, New York, 1993.

[57] B. Kerr. Niche Construction and Cognitive Evolution. Thought A Review Of Culture And Idea, 2(3):250–262, 2008.

[58] L. Kounios, J. Clune, K. Kouvaris, G. P. Wagner, M. Pavlicev, D. M. Weinreich, and R. A. Watson. Resolving the paradox of evolvability with learning theory: How evolution learns to improve evolvability on rugged fitness landscapes. ArXiv e-prints, 2016.

[59] K. Kouvaris, J. Clune, L. Kounios, M. Brede, and R. A. Watson. How evolution learns to generalise: Using the principles of learning theory to understand the evolution of developmental organisation. PLoS Computational Biology, 13(4):1–20, 2017.

[60] K. N. Laland, B. Matthews, and M. W. Feldman. An introduction to niche construction theory. Evolutionary Ecology, 30(2):191–202, 2016.

[61] K. N. Laland, K. Sterelny, J. Odling-Smee, W. Hoppitt, and T. Uller. Cause and effect in biology revisited: Is Mayr's proximate-ultimate dichotomy still useful? Science, 334(6062):1512–1516, 2011.

[62] K. N. Laland, T. Uller, M. W. Feldman, K. Sterelny, G. B. Müller, A. Moczek, E. Jablonka, and J. Odling-Smee. The extended evolutionary synthesis: its structure, assumptions and predictions. Proc. R. Soc. B, 282(1813):20151019, 2015.

[63] R. Lande. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. Journal of Evolutionary Biology, 22(7):1435–1446, 2009.

[64] R. B. Langerhans and T. J. DeWitt. Plasticity constrained: Over-generalized induction cues cause maladaptive phenotypes. Evolutionary Ecology Research, 4(6):857–870, 2002.

[65] J. Lehman and K. O. Stanley. Abandoning Objectives: Evolution Through the Search for Novelty Alone. Evolutionary Computation, 19(2):189–223, 2011.

[66] N. A. Levis and D. W. Pfennig. Evaluating 'Plasticity-First' Evolution in Nature: Key Criteria and Empirical Approaches. Trends in Ecology and Evolution, 31(7):563–574, 2016.

[67] M. Locatelli. A note on the Griewank test function. Journal of Global Optimization, 25(2):169–174, 2003.

[68] L. Loison. Canalization and genetic assimilation: Reassessing the radicality of the Waddingtonian concept of inheritance of acquired characters, 2019.

[69] K. Z. Lorenz. The evolution of behavior. Scientific American, 199(6):67–74, 1958.

[70] K. Z. Lorenz. The Foundations of Ethology. Springer Verlag, 1981.

[71] J. B. Losos, T. W. Schoener, and D. A. Spiller. Predator-induced behaviour shifts and natural selection in field-experimental lizard populations. Nature, 432(7016):505–508, 2004.

[72] J. Masel. Genetic assimilation can occur in the absence of selection for the assimilating phenotype, suggesting a role for the canalization heuristic. Journal of Evolutionary Biology, 17(5):1106–1110, 2004.

[73] G. Mayley. Guiding or Hiding: Explorations into the Effects of Learning on the Rate of Evolution. 4th European Conference on Artificial Life (ECAL97), pages 135–144, 1987.

[74] G. Mayley. Landscapes, Learning Costs, and Genetic Assimilation. Evolutionary Computation, 4(3):213–234, 1996.

[75] E. Mayr. Cause and Effect in Biology. Science, 134(3489):1501–1506, 1961.

[76] F. Menczer and R. K. Belew. Evolving sensors in environments of controlled complexity. Artificial life IV. Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems, pages 210–221, 1994.

[77] F. Mery and J. G. Burns. Behavioural plasticity: An interaction between evolution and experience. Evolutionary Ecology, 24(3):571–583, 2010.

[78] F. Mery and T. J. Kawecki. The Effect of Learning on Experimental Evolution of Resource Preference in Drosophila Melanogaster. Evolution, 58(4):757, 2004.

[79] R. Mills. How Micro-Evolution Can Guide Macro-Evolution. PhD thesis, University of Southampton, 2010.

[80] R. Mills and R. A. Watson. On Crossing Fitness Valleys with the Baldwin Effect. Proc. 10th Int. Conf. Simul Synthesis Living Systems., pages 493–499, 2006.

[81] A. Moczek, S. E. Sultan, S. Foster, I. Dworkin, H. F. Nijhout, E. Abouheif, and D. W. Pfennig. The role of developmental plasticity in evolutionary innovation. Proceedings of the Royal Society B: Biological Sciences, 278(June):2705–2713, 2011.

[82] T. Morgan, P. Whitehorn, G. C. Lye, and M. Vallejo-Marín. Floral Sonication is an Innate Behaviour in Bumblebees that can be Fine-Tuned with Experience in Manipulating Flowers. Journal of Insect Behavior, 29(2):233–241, 2016.

[83] T. J. Morgan, J. W. Suchow, and T. L. Griffiths. What the Baldwin Effect affects depends on the nature of plasticity. Cognition, 197(January 2019):104165, 2020.

[84] C. J. Murren, J. R. Auld, H. Callahan, C. K. Ghalambor, C. A. Handelsman, M. A. Heskel, J. G. Kingsolver, H. J. Maclean, J. Masel, H. Maughan, D. W. Pfennig, R. A. Relyea, S. Seiter, and U. K. Steiner. Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. Heredity, 115(4):293–301, 2015.

[85] F. Muth, D. R. Papaj, and A. S. Leonard. Colour learning when foraging for nectar and pollen: Bees learn two colours at once. Biology Letters, 11(10), 2015.

[86] K. J. Niklas. Morphological evolution through complex domains of fitness. Proceedings of the National Academy of Sciences of the United States of America, 91(15):6772–6779, 1994.

[87] S. Nolfi. How Learning and Evolution Interact: The Case of a Learning Task which Differs from the Evolutionary Task. Adaptive Behavior, pages 231–236, 1997.

[88] E. Page, K. D. Waddington, M. K. Fondrk, and G. J. Hunt. Genetic determinants of honey bee foraging behaviour. Animal Behaviour, 50(6):1617–1625, 1995.

[89] R. E. Page, G. E. Robinson, M. K. Fondrk, and M. E. Nasr. Effects of worker genotypic diversity on honey bee colony development and behavior (Apis mellifera L.). Behavioral Ecology and Sociobiology, 36(6):387–396, 1995.

[90] D. R. Papaj. Optimizing learning and its effect on evolutionary change in behavior. In Chapter, volume chapter SI, chapter 6, pages 133–153. The University of Chicago Press, 1994.

[91] J. Peat and D. Goulson. Effects of experience and weather on foraging rate and pollen versus nectar collection in the bumblebee, Bombus terrestris. Behavioral Ecology and Sociobiology, 58(2):152–156, 2005.

[92] F. B. Pereira and E. Costa. The influence of learning in the evolution of busy beavers. In E. J. W. Boers, editor, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 2037, pages 421–431, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

[93] F. B. Pereira, P. Machado, E. Costa, and A. Cardoso. Graph based crossover: A case study with the Busy Beaver problem. Proceedings of the Genetic and Evolutionary Computation Conference GECCO-1999, II:1149–1155, 1999.

[94] M. Pigliucci, C. J. Murren, and C. D. Schlichting. Phenotypic plasticity and evolution by genetic assimilation. Journal of Experimental Biology, 209(12):2362–2367, 2006.

[95] A. C. Pontes, R. B. Mobley, C. Ofria, C. Adami, and F. C. Dyer. The evolutionary origin of associative learning. American Naturalist, 195(1):E1–E19, 2020.

[96] T. D. Price, A. Qvarnstrom, and D. E. Irwin. The role of phenotypic plasticity in driving genetic evolution. Proceedings of the Royal Society B: Biological Sciences, 270(1523):1433–1440, 2003.

[97] A. Rago, K. Kouvaris, T. Uller, and R. A. Watson. How adaptive plasticity evolves when selected against. PLoS Computational Biology, 15(3):1–20, 2019.

[98] N. E. Raine, T. Ings, A. Dornhaus, N. Saleh, and L. Chittka. Adaptation, Genetic Drift, Pleiotropy, and History in the Evolution of Bee Foraging Behavior. Advances in the Study of Behavior, 36(06):305–354, 2006.

[99] S. C. P. Renn and M. E. Schumer. Genetic accommodation and behavioural evolution: insights from genomic studies. Animal Behaviour, 85(5):1012–1022, 2013.

[100] D. A. Roff and D. J. Fairbairn. A test of the hypothesis that correlation selection generates genetic correlations. Evolution, pages 2953–2960, 2012.

[101] N. Saito, S. Ishihara, and K. Kaneko. Baldwin effect under multipeaked fitness landscapes: Phenotypic fluctuation. American Physical Society, 052701:3–8, 2013.

[102] M. Santos, E. Szathmáry, and J. F. Fontanari. Phenotypic plasticity, the Baldwin effect, and the speeding up of evolution: The computational roots of an illusion. Journal of Theoretical Biology, 371:127–136, 2015.

[103] C. D. Schlichting and M. A. Wund. Phenotypic plasticity and epigenetic marking: An assessment of evidence for genetic accommodation. Evolution, 68(3):656–672, 2014.

[104] H. Schwabl. Yolk is a source of maternal testosterone for developing birds. Neurobiology, 90(December):11446–11450, 1993.

[105] M. L. Siegal and A. Bergman. Waddington's canalization revisited: Developmental stability and evolution. Proceedings of the National Academy of Sciences, 99(16):10528–10532, 2002.

[106] A. Sih, A. Bell, and J. C. Johnson. Behavioral syndromes: An ecological and evolutionary overview. Trends in Ecology and Evolution, 19(7):372–378, 2004.

[107] G. G. Simpson. The Baldwin Effect. Evolution, 7(2):110, 1953.

[108] E. J. Slijper. Biologic anatomical investigations on the bipedal gait and upright posture in mammals-With special reference to a little goat born without forelegs II. Proceedings of the Koninklijke Nederlandse Akademie Van Wetenschappen, 45:407–415, 1942.

[109] E. C. Snell-Rood. An overview of the evolutionary causes and consequences of behavioural plasticity. Animal Behaviour, 85(5):1004–1011, 2013.

[110] E. C. Snell-Rood and M. K. Steck. Behaviour shapes environmental variation and selection on learning and plasticity: review of mechanisms and implications. Animal Behaviour, 147:147–156, 2019.

[111] J. A. Stamps and W. E. Frankenhuis. Bayesian Models of Development. Trends in Ecology and Evolution, 31(4):260–268, 2016.

[112] S. E. Sultan. Phenotypic plasticity and plant adaptation. Acta Botanica Neederlandica, 44(December):363–383, 1995.

[113] S. E. Sultan. Developmental plasticity: re-conceiving the genotype. Interface Focus, 7(5), 10 2017.

[114] S. E. Sultan and F. Bezzaz. Phenotypic Plasticity in Polygonum persicaria . I . Diversity and Uniformity in Genotypic Norms of Reaction to Light. Society for the Study of Evolution, 47(4):1009–1031, 2017.

[115] S. Tebbich, M. Taborsky, B. Fessl, and D. Blomqvist. Do woodpecker finches acquire tool-use by social learning? Proceedings of the Royal Society B: Biological Sciences, 268(1482):2189–2193, 2001.

[116] A. J. Tierney. The evolution of learned and innate behavior: Contributions from genetics and neurobiology to a theory of behavioral evolution. Animal Learning & Behavior, 14(4):339–348, 1986.

[117] P. D. Turney. Myths and Legends of the Baldwin Effect. Evolution, 133(3350):8, 2002.

[118] M. N. Verzijden, C. ten Cate, M. R. Servedio, G. M. Kozak, J. W. Boughman, and E. I. Svensson. The impact of learning on sexual selection and speciation. Trends in Ecology and Evolution, 27(9):511–519, 2012.

[119] C. H. Waddington. Genetic assimilation of an acquired character. Evolution, 7(2):118–126, 1953.

[120] C. H. Waddington. A Catastrophe Theory of Evolution. Ann. N.Y. Acad. Sci., 1(231):32–41, 1974.

[121] A. Wagner. Does Evolutionary Plasticity Evolve? Evolution, 50(3):1008, 1996.

[122] R. A. Watson, C. L. Buckley, R. Mills, and A. Davies. Associative memory in gene regulation networks. Artificial Life XII. Proceedings of the 12th International Conference on the Synthesis and Simulation of Living Systems, pages 194–202, 2010.

[123] R. A. Watson, R. Mills, and C. L. Buckley. Transformations in the scale of behavior and the global optimization of constraints in adaptive networks. Adaptive Behavior, 19(4):227–249, 2011.

[124] R. A. Watson, R. Mills, and C. L. Buckley. Transformations in the Scale of Behaviour and the Global Optimisation of Constraints in Adaptive Networks. Adaptive Behavior, 19(4):227–249, 2011.

[125] R. A. Watson and J. B. Pollack. Modular Interdependency in Complex Dynamical Systems. Artificial life, 4(11):445–457, 2005.

[126] R. A. Watson, G. P. Wagner, M. Pavlicev, D. M. Weinreich, and R. Mills. The evolution of phenotypic correlations and "developmental memory". Evolution, 68(4):1124–1138, 2014.

[127] J. J. Welch. What's wrong with evolutionary biology? Biology & Philosophy, 32(2):263–279, 2017.

[128] S. A. West, R. M. Fisher, A. Gardner, and E. T. Kiers. Major evolutionary transitions in individuality. Proceedings of the National Academy of Sciences, 112(33):201421402, 2015.

[129] M. J. West-Eberhard. Developmental Plasticity and Evolution. Oxford University Press, 2003.

[130] M. J. West-Eberhard. Phenotypic accommodation: adaptive innovation due to developmental plasticity. Journal of Experimental Zoology, 3048(September):610–618, 2005.

[131] D. F. Westneat, L. J. Potts, K. L. Sasser, and J. D. Shaffer. Causes and Consequences of Phenotypic Plasticity in Complex Environments. Trends in Ecology and Evolution, 34(6):555–568, 2019.

[132] R. Woltereck. Weitere experimentelle Unter-suchungen uber Artveranderung, speziell uber das Wesen quantitativer Artunterschiede bei Daphniden. Verh. D. Tsch. Zool. Ges., 1909:110–172, 1909.

[133] K. J. Zollman and R. Smead. Plasticity and language: An example of the Baldwin effect? Philosophical Studies, 147(1):7–21, 2009.