Baldwinian Accounts of Language Evolution

Hajime Yamauchi

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy

to

Theoretical and Applied Linguistics,
College of Humanities & Social Science School of Philosophy,
Psychology & Language Sciences,
University of Edinburgh

August 2004

Abstract

Since Hinton & Nowlan published their seminal paper (Hinton & Nowlan 1987), the neglected evolutionary process of the Baldwin effect has been widely acknowledged. Especially in the field of language evolution, the Baldwin effect (Baldwin 1896*d*, Simpson 1953) has been expected to salvage the long-lasting deadlocked situation of modern linguistics: *i.e.*, it may shed light on the relationship between environment and innateness in the formation of language.

However, as intense research of this evolutionary theory goes on, certain robust difficulties have become apparent. One example is genotype-phenotype correlation. By computer simulations, both Yamauchi (1999, 2001) and Mayley (1996*b*) show that for the Baldwin effect to work legitimately, correlation between genotypes and phenotypes is the most essential underpinning. This is due to the fact that this type of the Baldwin effect adopts as its core mechanism Waddington's (1975) "genetic assimilation". In this mechanism, phenocopies have to be genetically closer to the innately predisposed genotype. Unfortunately this is an overly naïive assumption for the theory of language evolution. As a highly complex cognitive ability, the possibility that this type of genotype-phenotype correlation exists in the domain of linguistic ability is vanishingly small.

In this thesis, we develop a new type of mechanism, called "*Baldwinian Niche Construction* (**BNC**), that has a rich explanatory power and can potentially overcome this bewildering problem of the Baldwin effect. BNC is based on the theory of niche construction that has been developed by Odling-Smee *et al.* (2003). The incorporation of the theory into the Baldwin effect was first suggested by Deacon (1997) and briefly introduced by Godfrey-Smith (2003). However, its formulation is yet incomplete.

In the thesis, first, we review the studies of the Baldwin effect in both biology and the study of language evolution. Then the theory of BNC is more rigorously developed. Linguistic communication has an intrinsic property that is fundamentally described in the theory of niche construction. This naturally leads us to the theoretical necessity of BNC in language evolution. By creating a new linguistic niche, learning discloses a previously hidden genetic variance on which the Baldwin

'canalizing' effect can take place. It requires no genetic modification in a given genepool. There is even no need that genes responsible for learning occupy the same loci as genes for the innate linguistic knowledge. These and other aspects of BNC are presented with some results from computer simulations.

Acknowledgements

I would like to express my gratitude to all those who made it possible for me to complete this thesis.

I am deeply indebted to my supervisor Prof. Dr. James R. Hurford whose help, stimulating suggestions, and encouragement helped me during all the time of research for, and writing of this thesis. I have furthermore to thank Dr. Simon Kirby, who introduced me to the more technical part of the subject with great patience and pedagogical skills.

My colleagues from the Language Evolution & Computation Research Unit supported me in my research work. I want to thank them, in particular Dr. Andrew Smith, Dr. Kenny Smith, Dr. Henry Brighton, Jelle Zuidema, Dan Dediu, for all their help, support, interest and valuable hints. A number of Dan Dediu's accute comments were particularly informative. I am especially obliged to Anna Parker who looked closely at the thesis for English style and grammar, correcting both and offering suggestions for improvement. My sincere appreciation goes to Dr. Takashi Hashimoto. He was always considerate, and encouraged me so much. Special thanks go to Joseph Poulshock who has given me helpful suggestions on portions of this work at various stages of its production.

I also want to thank Takeshi Ishihara who was of great help at difficult times. I am very grateful for the love and support of Yoko Takahashi.

I owe Prof. Teruo Inoue so much that I can describe here only a very small portion. Prof. Inoue not only began my interest in language, but throughout my academic life, he has always been very supportive, and I will be grateful to him forever. I wish to thank Prof. Yukio Otsu; I became acquainted with formal linguistics in the lecture given by him. Also, during my first year in abroad, my awkward feelings of "being abroad" were minimized by his warmth. I was very much impressed by Prof. Philip Lieberman's course on evolutionary linguistics. It was his lecture that made me consider going into this field. My thanks go to Prof. Neil Smith whose always logical, calm attitude stimulated me so much. I am very grateful to Prof. Terrence W. Deacon whose influential book literally made my

mind up to finally go into this field, and through the book I learnt so much. This thesis truly stands on his brilliant concept of the Baldwin effect.

Last, but by no means least, I would like to give my special thanks to my parents, Etsuko and Osamu whose patient love enabled me to complete this work.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

For decades, the innate capacity of language acquisition has been one of the central issues of the study of language. How heavily does language acquisition rely on innate linguistic properties? This question, often called the "*nature–nurture*" debate, brings about endless debates involving linguists, psychologists, and even computer scientists. Indeed, it would not be too much to say that this has been the most fundamental axis of the different camps in modern linguistics. It would even be possible to summarize why schools in linguistics have been so seriously segregated from the perspective of this nature–nurture debate. Having said that, a number of phenomena that occur during language acquisition are quite puzzling when one tries to determine what parts of language acquisition are innate or attributed to postnatal learning. Mainly from its methodological restrictions, most of the studies in this area have dealt with the problem from more or less synchronic points of view. Consequently, currently available linguistic or psychological data are vital for formulations of linguistic theories. Thus, although this type of study may shed light on the nature of language acquisition, it might not be capable of providing us with an account of the origin of such a complex aspect of language acquisition even in a synchronic sense. This is of particular interest regarding the recent agreement of the nature–nurture problem; an intensive array of studies has gradually revealed that this twofold structure of language acquisition never appears as a clear dichotomy. Rather, the intriguing interaction between innate and learnt properties of language seems to require a new avenue of linguistic studies.

Language diversity also poses a similar complication. Although there is no agreed total, most reference books give a figure of 5,000 to 6,000 for the number of languages. The world's languages have more or less equal communicative powers; there is no language which is more 'primitive' than other languages, no matter how 'primitive' its speakers are in technological terms. However, the structure of each language that contributes to such an expressive power may vary quite notably. This is well reflected in the study of language; linguists have dedicated 40 years to

finding the underlying commonality of the world's languages, but the results are still widely open to interpretation. While we are struggling to find such universalities that define 'language' as a whole, it has been an undeniable fact that no child has any problem acquiring any natural language. This problem also demands us to search beyond the current linguistic paradigm.

To speculate on these problems, it will be fruitful to consider the obstacles in our understanding of the two problems noted above. Our current studies of these fields have put great emphasis on individuals. This attitude may provide a correct avenue to pursue the questions to a certain extent. After all linguistic activity can be attributed to human cognitive activities. However, language has a different aspect which is no less important than the former, namely a social and dynamic aspect. We are in a linguistic arena where all types of linguistic activities take place. Thus to understand the previous questions, it may be important to shift from the current linguistic emphasis on individuals to study with a more *populational* perspective. Such a populational study will naturally lead us to speculate on more diachronic and dynamic aspects of language acquisition. Then this would be a new avenue for the previous problems.

When we combine this populational view of language with the consideration of language acquisition in the context of an evolutionary perspective, an interesting point emerges. Recent surveys in the field of computational simulations reincarnate a more-than-100-years-old argument in evolutionary study. In 1896, an American psychologist James M. Baldwin (1896*d*) proposed "*a new factor in evolution*". He assumed that if an individual is capable of acquiring an adaptive behavior postnatally, addition of such a learning process in the context of the evolutionary search potentially changes the profile of populational evolution; learning paves the path of the evolutionary search so that the burden of the evolutionary search is eased. In addition, this special synergy of learning and evolutionary search has a further effect; a phenomenon in which "*a behavior that was once learned may eventually become instinctive*" (Turney *et al.* 1996).

This learning-guided evolutionary scenario, known as *the Baldwin effect* (Simpson 1953), possibly provides a strikingly attractive perspective to the nature–nurture problem in linguistics. It has been attested by a number of computer simulations (*e.g.*, Hinton & Nowlan 1987) in the field of computer science that if an environment surrounding a population is prone to shift to a new environment, some learning is adaptive. If those environments do not share any commonality, an individual who relies on learning for every aspect of the behavior will be the most adaptive. However, if those environments hold some universality, an individual who has partially genetically-predisposed and partially learned behavior will be the most adaptive;

for example, the predisposed part of the behavior covers the commonality and the learned part of the behavior covers the differences (this point is discussed in Chapter 2).

The Baldwin effect in linguistics may shed significant light on some long-standing problems. For example, preliminary studies suggest that language evolution is out of the scope of natural selection mainly because of its dysfunctional nature. For those researchers, language evolution is a consequence of exaptation (Gould & Vrba 1982) or a big leap in evolution (*see* Chomsky 1972, Chomsky 1982*a*, Chomsky 1982*b*, Newmeyer 2000, Piatelli-Palmarini 1989). This no-intermediate-stages scenario would be, however, explicable by natural selection when it is guided by learning since learning can smooth the intermediate landscape. Subsequently, it has been a popular idea that the Baldwin effect is a crucial factor in the evolution of language (*e.g.*, Waddington 1975, Pinker & Bloom 1990, Briscoe 1997, Briscoe 2002*a*, Briscoe 2002*b*, Turkel 2002).

Moreover, as the prominent linguists Steven Pinker and Paul Bloom (1990) concisely indicated (*see* Section 3.1.2), the Baldwin effect would greatly contribute to the apparent problem of language evolution; if language evolution is not saltational, but gradual, how would qualitative discrepancies among individuals regarding their communicability be circumvented? In normal evolution, such differences are directly connected to differences of fitness among the individuals. However, in a communication system, such differences would mostly work as obstacles in communications; it is more or less meaningless if someone has a 'better' envelope of communication; it cannot be used because others would not comprehend. Pinker & Bloom suggested that the Baldwin effect would make it evolvable.

In summary, the Baldwin effect is particularly appealing because of the following three reasons:

1. It may provide a new perspective to tackle the nature–nurture problem: Because the Baldwin effect deals with interactions of learning and innateness in an evolutionary perspective, it is expected to provide a new avenue to consider how the "*Language Acquisition Device*" (**LAD**, Chomsky 1981) has been formed, and what the possible quality of the device is.
2. It may provide a natural Darwinian account for language evolution: It is an especially popular idea among linguists that language evolution is somehow saltational. This leads them to conclude Darwinian theories are 'incompetent' for accounting for language evolution.

3. It nicely connects learning –a process at the individual level with evolution –a process in the population level: Given the facts that learning is one of the most crucial aspects of language, and its inputs come from a previous generation, the Baldwin effect may be able to unite the cultural evolutionary aspect of language and its phylogenetic aspect.

Therefore, in the study of language evolution, the Baldwin effect has been by and large welcomed for the above reasons. Including studies which casually refer to the Baldwin effect in the non-focal part of their argument, the number of works which adopt the theory is non-trivial. However, it is also true that the Baldwin effect has been somehow treated as some sort of '*Deus ex Machina*' in theories of language evolution; the majority of such works, even including Pinker & Bloom, have not paid serious attention to the theoretical validity of the effect[1].

Unfortunately, things are not so rosy. First we have to notice that even if the Baldwin effect provides a suggestive view for the riddle of both the nature–nurture problem and language evolution itself, the theory of the Baldwin effect itself does not provide a desirable theory that is tenable to rigorous scientific examinations.

Even worse, the Baldwin effect itself is controversial. George G. Simpson, the paleontologist and the one of the founders of the Modern Synthesis, to start with, was skeptical about the concept even though it was he who promoted it as a modern evolutionary theory. Although in the late 1980's, the Baldwin effect was 'rediscovered' by computer scientists and has been applauded for a decade or so, recently, reconsideration of the theory seems to be a more definite trend which mainly arose from developmental biology.

As will be discussed in Chapter 4, as intense research of this evolutionary theory goes on, certain robust difficulties have become apparent. One example is the genotype-phenotype correlation; for the Baldwin effect to work legitimately, correlation between genotypes and phenotypes is the most essential underpinning. This is due to the fact that this type of the Baldwin effect adopts as its core mechanism a comparatively optimistic genetic assumption.

This thesis is not primarily directed to address how these perspectives indeed influence our understandings of language and its evolution. Nor is this a place in which we develop a theory of language evolution. The gap between what the Baldwin effect can provide and a desirable evolutionary theory which is falsifiable and tenable to rigorous theoretical examinations is still large. Indeed, given the above problem of genotype-phenotype correlation, the current mechanism of the

[1]However, there are some rare exceptions (*e.g.*, Deacon 1997, Dor & Jablonka 2000, Dor & Jablonka 2001).

Baldwin effect seems to be highly implausible. Instead, in this thesis, we propose a new type of mechanism which enables the Baldwin effect to emerge reliably even under genetically complex circumstances. The mechanism is equipped with a rich explanatory power and shows an especially high compatibility to language evolution. As such we believe that it would shed significant light on our understandings of language evolution.

The structure of this thesis is as follows: In the rest of this chapter, we discuss some basics about the concept of innateness. As the Baldwin effect is thought of a process of increasing innate attribution to learnt behavior, consideration of innateness itself is important. In the next chapter, we review some of the basic understandings of the Baldwin effect with literature reviews. This includes literatures in both evolutionary biology and computer science. Then, some studies of language evolution that explicitly adopt the Baldwin effect will be examined in Chapter 3.

In Chapter 4, we will examine the concept from a more critical point of view. Chapter 5 gives a view of a recent development of a constructive approach in evolutionary study, namely Niche Construction. Given the discussion, in Chapter 6, we propose a new mechanism of the Baldwin effect, called Baldwinian Niche Construction. Also, some important contributions of the mechanism will be considered in the chapter. Baldwinian niche construction in language evolution is considered in Chapter 7. Chapter 8 provides a basic idea of Chomsky's Principles and Parameters theory which is often used for the linguistic acquisition mechanism in computer simulations. In Chapter 9, we conduct several computer simulations including some replications of previous studies. Given the result, Chapter 10 discusses understandings of our study. Finally, the appendix provides some basic concepts of neutrality which is considered to be important for Baldwinian Niche Construction.

1.1 What is Innateness?

1.1.1 Innateness as Phenomenon

Innateness is one of the concepts the term "nature" entails. As *nature* and *nurture* are metaphorical terms, they entail similar but different notions. "Nature" generally refers to: 1.*instinctive* trait, 2.*innate* trait, 3.*inherited* trait, and 4.*genetic-base* trait. On the other hand, "nurture" often means: 1.*acquired* trait, 2.*learnt* trait, and 3.*environmentally-induced* trait (Table 1.1).

Baldwin himself considered the instinctive trait first in his mind, while current researchers generally consider the other three types of traits. Also, the evolutionist Conrad H. Waddington (1975) experimentally showed that environmentally-induced traits can change to a more-or-less fully genetic-based trait, while he theoretically promoted this idea in both acquired and learnt behavior (with instinctive and/or

	NATURE	NURTURE
TRAIT	instinctive innate inherited genetic-base	acquired learnt environmentally-induced

Table 1.1: Inventories of Nature–Nurture

innate behavior)[2]. Obviously, linguists pay much attention to innateness, but this does not necessarily mean that it is instinctive. When the Baldwin effect is considered in the context of language evolution, the most related traits will be innate and learnt ones.

Although this classification may clarify what is our real target when we are saying "innate" (say, it is neither genetic-based nor instinctive), the term innate itself is notoriously ambiguous; almost every argument that concerns innateness at any level is embarrassingly moot. It is quite common that once something is recognized as an innate property, later on, it is reevaluated that a large proportion of the property comes from an allegedly 'unrelated' behavior. A well-known example of early ethologists' optimistic claims of innateness in the 1940's and 1950's (*e.g.*, Lorenz and Tinbergen) and its subsequent refusal clearly depicts the point (*see* Lehrman 1953).

Consider, for example, the 'deprivation experiment' –the once most widely used technique to assess the possible innate ability of an organism. In this type of experiment, one typically sets a condition which is designed so that animals are raised in 'social isolation' to exclude inputs which allegedly contribute to form concerned behavior. In other words, it attempts to create a 'vacuum' condition. However, this experiment suffers apparent logical difficulties; first, the 'deprived' environment is still 'an environment'. And secondly, it is fundamentally inscrutable whether related inputs are still available under such an environment.

A clear example comes from Gottlieb (1971). It is known by ethologists that mallard ducklings and chicks can identify the maternal assembly call of their own species after hatching. Gottlieb found that this still holds even if they hatch in incubators where no maternal contact is available (thus it is a deprived condition). This suggests that mallard ducklings' preference of their species-specific parental call is innate.

[2] *see* Chapter 2.

However, a couple of years later, he found that if embryos are devocalized without interfering with other traits, ducklings could not distinguish the mallard parental call from chicken parental calls. This clearly shows that the preference of the mallard species-specific parental call hinges on hearing the 'contact-contentment' call that is produced by themselves in the shell.

Together with the blistering rejection of Skinnerian behaviorism, these sharp critiques of early ethology brought, as a consequence, fierce debates regarding the concept of innateness. And some are extremely doubtful about the innateness concept itself by claiming that such a concept emerges because our sense is deeply contaminated by 'folkbiology' (*e.g.*, Griffiths 2002).

These studies have been blowing a whistle to almost all fields that deal with any sort of informational, behavioral properties of living creature. However, it seems that not much debate has taken place on the study of the Baldwin effect, even though the concept has gathered high attentions; when one talks about the concept, the sharp dichotomy of nature–nurture is somehow brought back in: And it is largely unquestioned. It may be because debates revolving around possible explanatory adequacy of the concept are yet unsettled. Nevertheless, without sharing a common ground in terms of the definition of the innateness concept *within* the argument of the Baldwin effect, any theoretical attempt based on the concept will be fundamentally vacuous. In the following section, some basic understandings of the innateness concept will be examined and considered within the context of the Baldwin effect.

Innateness and Domain-Specificity

After all, everything is innate in some sense and to some extent. Ned Block brilliantly describes this in the following passage:

> No organism can learn without a mechanism that accomplishes this learning. Hence at least one learning mechanism must be innate (if only a mechanism for acquiring other learning mechanisms).
> (Block 1981, p. 279)

This seems to be an awfully banal statement, and apparently this can only serve as a 'grand theory' of innateness. However, it would be the only level where all researchers both pros and cons of nativism could meet; any higher level argument would be controversial. Unfortunately, however, often this type of grand theory has been wrongly incorporated into scientific studies.

Consider, for example, linguistic innateness. It has been often said that only human beings can acquire languages, and not other animals. If we accept the fact that what differentiates between human beings and other species is ultimately genes, it

means that we naturally admit that linguistic ability is somewhat innately rooted. From this 'grand fact', one starts at a wrong assumption; linguistic knowledge is 'innate' and to be equipped with as such an ability is a species-specific property. However this is obviously a farfetched conclusion; while there must be some relationship between this species-specific ability and innateness (as the above reason shows), a possible causal relationship is arbitrarily long. Then, it is clear that assuming that linguistic knowledge is innate based on species-specificity should be carefully considered. Similarly other forms of specificity-related properties are also easily confused with innateness; uniqueness or idiosyncrasies of linguistic knowledge are often treated as *ipso facto* the evidence of innateness. Elman *et al.* recount the following four different categories regarding the confusions around innateness and domain specificity (Elman *et al.* 1996, Bates *et al.* 1998):

1. Innateness and Species Specificity
2. Innateness and Domain Specificity
   (a) Behavioral Specificity
   (b) Representational Specificity
   (c) Specificity of Mental/Neural Processes
   (d) Genetic Specificity
3. Innateness and Localization
4. Innateness and Learnability

The description of 1 is already given above. 2 is subcategorized into four further different types.

The first is behavioral specificity and innateness. Although individual languages are (significantly, in some sense) different, such differences are fundamentally unlike other causally resembled animal abilities (*e.g.*, birdsong –learning in vocal channel, chess –a complex set of solutions to a game that only humans play, or music –rule-governed transitions in sound. Bates *et al.* 1998). Thus commonalities among languages are fundamentally different from commonalities between the linguistic system and other cognitive abilities. This fact is part of the evidence that such a linguistic ability could be domain specific. However, this does not directly support the fact that it is an innate ability. Bates *et al.* (1998) maintain that languages represent a class of solutions to a highly idiosyncratic problem of mapping a hyperdimensional meaning space onto a low-dimensional channel, and such a meaning space could be ubiquitous among human beings as experiences involved in forming such a meaning space would be shared by all normal members of the species. Besides, information processing required for this signal system may require specific

channels that are subject to universal constraints. If this is indeed the case, it would be a source of the domain specificity.

2b states that if a given problem is domain specific, behavior solving such a problem must be implemented by a set of domain specific mental/neural representations. This is regardless of whether such representations are innate or learnt.

2c is specificity of mental/neural process; it is known that the visual cortex of the cat contains strange neurons which only serve for lines at a specific orientation. This type of neuron is a highly peculiar one so that one might assume that it is innate. However, it is logically conceivable that such a neural process is still learnt. This view is strengthened by the fact that in neural network simulations, such peculiar structures continuously emerge in a multilayered neural network whenever extracting three-dimensional information mapped onto a two-dimensional image. This suggests that such structures could be postnatally constructed as they are required to process vital visual information in the animal's daily life. Elman *et al.* presume that this could be applicable to linguistic knowledge.

2d is also often confused with innateness. In our daily life, we often encounter various news on TV, newspapers, or magazines which report the 'discovery' of genes for *X*. However, careful reading soon reveals that most of such articles are simply vacuous; the putative relationship between a particular gene (or a set of genes) and a phenotypic trait is causally so fragile and there is no room for such a strong claim to be made. Very similar arguments are found in scientific research too; although they are often protected in logically more sophisticated arguments, there are a number of claims that suggest the existence of straightforward relationships between genes and phenotypic traits. Specific Language Impairment (**SLI**) has been treated as evidence of the strong genetic foundation of linguistic knowledge. As Bates *et al.* point out, however, without considering intervening levels, concluding such a genetic foundation of linguistic ability is next to meaningless; after all, any genetic disorder depriving one of one's ability to acquire languages can be used as evidence of the domain-specificity of linguistic ability. In the extreme case, like cerebral agenesis (*i.e.*, cases of genetic diseases where no functional brain develops above the level of brainstem), no language will emerge. But no one dares to claim that language has a genetic basis with this example.

Both 3 and 4 are in the same vein of the above discussion. These are not necessarily, by the very fact, evidence for innateness. These assure that debates revolving around the concept of innateness hold a highly sensitive assumptions.

Finally, domain-specificity also requires some caution when it is discussed in the context of evolution. It is almost a natural response to seek an evolutionary origin of domain-specificity when something is recognized as domain specific. However,

in the same way as the innate argument described above, often this ends up as a hunt for a fallacious origin of such a domain-specific property. What is important to consider is the possibility that it may be an consequence of 'exaptation' (Gould & Vrba 1982). Exaptation is also known as "preadaptation". As the name shows, it describes adoption of a trait which evolved for a functionally different demand. Karmiloff-Smith *et al.* state a possible relationship between exaptation and domain-specificity by citing a linguistic example:

> Having a vocal tract that makes a right-angle bend is very useful for language because it enriches the repertoire of sounds that can be produced. But did evolution create this angle 'for language'? Probably not, because it is also the result of upright bipedalism. However, it does have a domain-specific *consequence*. **We should not equate starting state and outcome when considering either evolution or ontogeny**. (Karmiloff-Smith *et al.* 1998, p. 590: original emphasis in italics, bold style added)

Innateness and Development

The above section considers the concept of innateness somewhat synchronically; innateness is often confused with domain-specific properties that are fundamentally irrelevant of innateness or not. This section, on the other hand, examines the relationship between innateness and development.

As in the previous example of mallard ducklings, the deprived method used by the early ethologists is based on the fallacious, optimistic assumption; if 'related experiences' are excluded, and still the target behavior is observable, then it can be judged as innate. The truth is, the definition of 'relatedness' was highly dubious.

In this regard, Johnson & Morton's (1991) taxonomy provides a useful insight. The taxonomy consists of different categories based on levels of interaction between genes and the environment. A simplified classification appears in Table 1.2. In the table, the term "innate" is confined to refer only to outcomes that arise as a consequence of interactions within a given organism. The second class, species-typical environment is a type of environment which all organism in the same species will experience in a normal developmental process. Johnson & Morton defines this as follows: "*. . . we will use the term primal to refer to any cognitive mechanism that results from the interaction between the animal (even before birth of or hatching) and any non-specific aspect of its species-typical environment*" (Johnson & Morton 1991, p. 10: original emphasis). In the case of the mallard duckling example shown above, it appears that mallard ducklings hear their own or their neighbor siblings prenatal, embryonic 'contact-contentment' calls before hatching. This environment

ENVIRONMENT	TERMS
Internal environment	INNATE
Species-typical environment (STE)	PRIMAL
Individual-specific environment (ISE)	LEARNING

Table 1.2: Levels of gene-environment interaction

is species specific, and individually non-specific; all young individuals of the species are normally expected to encounter it. In other words, primal environmental factors are those factors which are ubiquitously available for the given species.

This notion of species-typical environment (**STE**) and 'primal' is highly suggestive. In this view of innateness and related developmental processes, the reason that the early ethologists such as Lorenz or Tinbergen misunderstood (together with their methodological mistakes) was rooted to the confusion of these two independent classes; the notion of innateness should be restricted to outcomes of interactions which have taken place within organism-internal environments, such as molecular, cellular levels, while primal outputs are generated by the organisms' interactions with a class of external environments. Unfortunately, with their poor methodology and the rather simple assumption, the differences were neglected; superficially, outcomes of both levels of interactions are the same.

In the taxonomy, learning is located on the highest level of interaction. The related environment is called an individual environment since at this level, experienced environmental conditions will be different among individual organisms.

1.1.2 Innateness as Mechanism

Levels of Innateness

The above two sections show the phenomenal notion of innateness and other easily mixed concepts. In this section, in turn, the mechanistic property of innateness is briefly examined.

From the developmental perspective, innateness can be considered as constraints on developmental process. As in the above section, innateness is the product of interactions in organism-internal environments. The next question is, then, about the types of constraints in the process on the organism-internal level. Elman *et al.* (1996) propose the following three classifications:

1. Representational Constraints
2. Architectural Constraints
   (a) Basic Computing Units
   (b) Local Architecture
   (c) Global Architecture

3. Chronotopic Constraints

As the name shows, 1 refers to constraints on innate structuring of the mental/neural representations. Elman *et al.* assume that synaptic connectivity is the most likely candidate for this level of constraints.

2 can be divided into three further categories. 2a is considered to be the elements of architectural constraints which form a basic unit of computation. For example, basic rules on chemical reactions in synapses or excitatory/inhibitory properties of neuron can be classified in this category. In an example of a computer, this level would be equal to basic computation operators (such as AND/OR). 2b considers regional factors such as density of different cell types within layers, the number and thickness of such layers. This would be the processing unit of the computer; types of operators, and the way they are assembled determine what type of computation the processing unit is good at (say, CPU or FPU). The last and the highest level of architectural constraint is "Global Architecture". This includes afferent/efferent pathways whose connections integrate different local architectures.

Chronotopic constraints are the highest, and possibly indirect innate constraints. These are on the timing of developmental events, such as cell division taking place in neurogenesis, synaptic growth. Timings of these developments may not be directly hard-coded in genes, but such timings may produce significant constraints on the properties of various traits.

Innateness & Genetic Determinism

As in the discussion of innateness and domain-specificity, there is a great tendency to assume a straightforward causal relationship between an innate property and genetic attribution of such a property. This is often called "*genetic determinism*". Since genes, the unit of selection, are ultimately involved in evolution, this also non-trivially affects our way of thinking of innateness in the context of evolution; evolution is a process which fundamentally produces or fortifies innate attribution of a certain trait.

However, obviously various levels of causes are conceivable in mechanisms of innateness. Genes are one of such candidates; recent studies have found that significant numbers of innate properties are not solely due to gene(s), but cycles of developmental reactions are responsible (if not by STEs). Even though genes are somehow the 'masterminds' behind of these reactions, it would be logically implausible to call them "the cause" of a given innate property.

1.1.3 Learning, Plasticity, and Acquisition

In the study of language evolution, at least, the terms plasticity and learning have been interchangeably used. However, there are some important differences between

these two notions regarding researchers' recognitions. In linguistics, it is well known that some significant chronotopic constraint exists in language acquisition. This was hypothesized by Lenneberg (1967) –*The Critical Period Hypothesis*. It purports that language acquisition is a time-sensitive process; to be one's mother tongue, a language has to be learnt by a learner before arriving at puberty. This is similar to the concept of imprinting or crystallization in other animals, at least a superficial level, though the process will be much more complex and quite different mechanisms may be involved. The important point is that the concept of learning is not usually used as an ever-changing ability in the response to environmental drifts or other factors (this is only so in science). Rather, such an ability should be captured as part of the development which leads organisms to a certain (probably mildly) fixed phenotypic point. Beyond that point, no significant change would take place. Or, even so, it may not be driven by learning, but some other factors would be responsible for it.

Often such phenotypic endpoints are confined to within a certain range; regardless that it is directly controlled by genes or by some systematic effects, a possible range of phenotype after the learning process is statistically determined. In linguistics, this is considered as learnability of languages. Although it has not been empirically attested (because of both practical and ethical reasons), it is widely accepted that one can define a class of symbolic systems which corresponds to a set of learnable languages.

Also, plasticity is usually used in wider contexts than learning in biology. This is natural as the term plastic simply means the moldable and restorable property, while learning can only be used for some higher animates' acquiring knowledge (including behavioral knowledge). Thus when we discuss plasticity of some biological property, it could refer to anything from a cell level object such as a synapse to a more abstract object such as linguistic ability. Furthermore, plasticity is more often used for physical objects while learning is used for mnemonic, and/or behavioral properties almost exclusively. Thus those two terms can be differently used in the same context. A good example is often found in neuroscience. For instance, in the study of memory, continuous updates of memory (*i.e.*, learning) are usually attributed to adjustments of plastic synapses. These are considered to be related to consciousness and/or spontaneousness of the organism involved in the modification. On this point of view, learning is generally attributed to a conscious, and spontaneous modification, while plasticity would be more often considered in automatic, imperative physiological or neurological responses. However, especially in the case of first language acquisition, learning is likely to be unconscious, and non-spontaneous. This is a partial reason that linguistics has a long history of using

the term acquisition in lieu of the term learning (though it is far from an exclusive use).

The hint of the reason that these have been interchangeably used in the context of the Baldwin effect is found in table 1.1. As discussed above, the Baldwin effect is considered in different contexts. Like Baldwin himself, if one considers instinctive behavior in his mind, the term learning is more suitable. Instead, like Waddington, if he thinks of acquired physical traits, plasticity would be appropriate, especially when he is dealing with physiologically lower level traits. In the same vein, acquisition would be used in the study of language evolution.

Keeping these differences in mind, for practical reasons, in this thesis these are interchangeably used unless otherwise noted.

1.2 Innateness and the Baldwin Effect

As well reflected in Turney *et al.*'s (1996)'s comment, the Baldwin effect has been mostly considered within the context of increasing innate attribution of a learnt trait. The trouble is, if a once-learnt trait 'ineluctably' appears during epigenetic development, there is no guarantee that it is due to increase of innate attribution; it should be clear from the above discussions that a stable emergence of such a trait may well be the result of STE; the consistent reaction may be derived just because certain environmental factors are ubiquitous for all individuals in a given population, thus all inputs necessary for learning of a particular trait are the same. Under this circumstance, even highly plastic individuals would develop the same (or highly similar) trait as more innately predisposed individuals do.

This becomes especially important when a new mechanism of the Baldwin effect (called Baldwinian Niche Construction) is concerned; part of a given environment is formed through organisms' own learnt activity, and it is non-genetically, but culturally inherited over generations. As such, it may well become an STE.

Secondly, if a given trait is found to be indeed innately predisposed, the cause of such an 'innate attribution' is often blindly related to 'genetically hard-coded' property (*i.e.*, genetic determinism). This is a rather salient tendency in the studies of the Baldwin effect. As we will see in the next chapter (also in Chapter 4), an ambiguous usage of the term "*genetic assimilation*" would be mostly responsible for this. However, various levels of causes and agents are still conceivable for a given innately predisposed development.

Thus, the issue revolving around the Baldwin effect and innateness is two-fold. First, the causes of increasing ineluctability of a once-learnt trait in a developmental stage are pluralistic; it could be attributed to either STEs or innately more

predisposed property. Secondly, even if the process indeed increases the innate attribution, the causes are also pluralistic; there are various levels conceivable between the innate attribution and genetic contributions.

While increase of ineluctability through STEs should be excluded as a case of the Baldwin effect[3], different causes of innate attribution should be considered within the context of the Baldwin effect. Therefore, in this thesis, the Baldwin effect is considered in the context of innateness but not in that of STEs. Furthermore, given the above discussion, we accept different levels of innateness. This is especially important as in Chapter 9, all simulations adopt a simple, unrealistic model of 'gene-innate' relationship; such a model is used purely because it enables causally clear interactions between learning and evolution in simulations, and not because it reflects our view of innateness.

[3]But this will be revised in Chapter 10.

Chapter 2

The Baldwin Effect

2.1 A Brief History

The basic concept of the Baldwin effect was originally expressed by the American psychologist James M. Baldwin (1896*d*) more than 100 years ago. From our sense, the formulation of his concept was crude and somewhat vague, as at that time Mendel's work on inheritance was not even rediscovered. As his primary interest was directed to what he called "*social heredity*" (*see* Section 5.4) but not the Baldwin effect itself, for the next several years his study of the effect did not take off very much.

About 60 years later, Simpson (1953) formally termed Baldwin's idea "the Baldwin effect" within the context of the modern synthesis. He reformulated the effect more clearly and discussed possible examples of the effect. However, the overall tone was, as we will discuss in a later chapter, mostly dismissive.

Waddington had independently studied a similar concept called the "*canalization*" process and "*genetic assimilation*". Although the precise natures of theses concepts are slightly different from the Baldwin effect itself, his studies have greatly contributed to enhance theoretical adequacies of the Baldwin effect.

In the late 1980's when computational resources finally became handy for researchers, Geoffrey E. Hinton and Steven J. Nowlan (1987) conducted a simple simulation, and convincingly demonstrated the viability of the effect. This work, together with the 100 years milestone of Baldwin's original work, caused a new epoch in studies of the Baldwin effect and Baldwinian accounts of evolution suddenly flourished.

This short history of the Baldwin effect can also be considered as the history of clarification of the concept itself. For example, Simpson's pithy classification described below enables us to consider what the Baldwin effect is in a modern sense, while Hinton & Nowlan successfully depicts a picture of the potential implication of the Baldwin effect. These works have provided their own important contributions to understanding the Baldwin effect in the study of evolution.

Consequently, a number of studies in this field have been sparked by these studies in the 1990's. In this chapter, we will examine the Baldwin effect in some depth. First, we will look at the conventional notion of the Baldwin effect. Then literature from computational studies will be examined. In the following chapter, literatures on language evolution that appeal to the Baldwin effect will be considered.

2.2 The Basic Formulation

While the Baldwin effect has been intensively discussed both directly (*e.g.*, as an explanandum) and indirectly (*e.g.*, as an explanans), it has been criticized that the definition of the effect has never been fixed and is expressed in various ways. The greatest common factor among such studies would be the following classification described in Simpson (1953):

Stage 1 In an environment, a new condition has emerged. No individuals in the population have yet undergone an adaptation.

Stage 2 In the population, individuals interact with the environment in such a way as to systematically produce behavioral, physiological, or structural modifications that are not hereditary as such but that are advantageous for survival, *i.e.*, are adaptive for the individuals having them.

Stage 3 In the population, genetic factors producing hereditary characteristics similar to the individual modification referred to in (2), or having the same sorts of adaptive advantages occur. Such genetic factors are favored by natural selection and tend to spread in the population over the course of generations. The net result is that adaptation that was originally individual and non-hereditary becomes hereditary.

While the above description properly characterizes the basic aspect of the Baldwin effect, from this it is somewhat hard to realize that why the Baldwin effect is an intriguing evolutionary process. Studies in this field have revealed that two possible evolutionary impacts of the Baldwin effect are conceivable –the *expediting* effect and the *canalizing* effect. In the following sections, these two effects are described.

2.3 The Expediting Effect

It has been argued that evolution can be understood as '*hillclimbing*'. The gradual evolution of a certain trait increments its adaptive fitness gradually. Thus, if one individual possesses only 5% of our current visual system, it is better than no visual at all. If a descendant has 10% of our visual system, it certainly has better fitness than the ancestor with 5% does. Therefore, every small improvement of the current trait provides a better possibility of survival for an individual that possesses the

trait. In this case, natural selection provides quite a robust solution that steadily propels the individuals from the base to the summit.

However, it is logically conceivable that 5% or 10% of a system has no adaptive advantage compared with 0%, but only the 100% system can increase fitness. Also, in a dynamic environment, it is possible that one trait which had provided a good fitness in an old environment would never work well in a new environment. In these cases, natural selection could not provide a good evolutionary solution. In the former case, natural selection merely searches for the spiky summit randomly. In the latter, it has to start searching for the new summit after every single environmental shift. In extreme cases, natural selection never finds the summit in both situations.

In computer simulations that use the techniques of biological evolution (*i.e.*, the Darwinian evolutionary mechanism) –"*Genetic Algorithms*" (**GAs** Holland 1975), often this hillclimbing process is visually considered in a schematic landscape called a "fitness landscape". A fitness landscape is a visually described collective representation of the fitness values that all genotypes can take. It is often represented in a three-dimensional graph where the x- and y-axes represent a given genotype's position and the z-axis corresponds to the fitness value of the given genotype. Since any one individual has a unique genotype, it occupies a corresponding position on the landscape. Recombinations (*i.e.*, crossovers) and mutations are considered as the random redistribution process of such occupied positions. If a given fitness landscape has a unimodal mountain like single peak (called a "*Fujiyama*" landscape), evolution by the hillclimbing process efficiently finds the summit; it is hillclimbing, since natural selection favors individuals who occupy higher positions on the landscape than others.

If a landscape is spiky, however, the hillclimbing process may hardly take place; the random redistribution process cannot be followed by an efficient selection process since the target positions (through the redistributions) may well have no differences in their height (hence, fitness values).

This is where learning (or plasticity) becomes efficient. If an individual possesses phenotypic plasticity or an ability to learn, the individual can obtain better fitness within its lifetime since such abilities enable the individual to override what she genetically inherits when such an inherited property is proven to be unfit for a given environment. For simplicity, suppose that 'learning' attempts exactly the same task as the genetic search; both evolutionary search and learning share the same fitness landscape[1]. Since the pace of search by learning is faster than natural selection,

[1]As it becomes apparent in later, this exact mapping between genetic search and phenotypic search is the most crucial assumption in the currently dominant mechanism of the Baldwin effect (*see* Mayley 1996*b*).

the individual that possesses plasticity can take advantage of it to search for better solutions in the cases stated above. Then those individuals that reach the spiky summit within their life succeed in proliferating over the others that do not find the summit. As this is a selective process just as natural selection is, but happens within the ontogenetic span, Baldwin (1896*d*) named it "*Organic Selection*".

Organic selection enables individuals to be reproductive under the otherwise non-prolific environment. However, it is natural to consider that there are individual variations even in this ontogenetic search; some individuals are good at this type of search while others may not be. For our convenience, let us consider the maximum amount of the learning range is equal for all individuals. In other words, all individuals' learning ability is qualitatively the same. In this case, the only difference among the individuals is the distance from the summit on the fitness landscape; for some individuals, to find the summit they may travel quite a lot, while for others short travels would be enough.

Often learning (or retaining plasticity) is a costly option. While several types of costs can be considered in learning, for the time being let us only consider a cost created by the process itself. That is, by learning something individuals consume some energy which is otherwise not incarnated. In the scheme of the fitness landscape, this type of cost is equal to the traveling distance of individuals on the landscape during their learning. If the cost of learning is (negatively) reflected on their fitness, those who find the summit without much learning obviously can gain better fitness. Since such individuals typically have genotypes which are relatively close to the summit, this effectively makes the spiky fitness landscape a more smooth landscape. In other words, the combination of evolutionary search and learning smooths a spiky fitness landscape.

This "*smoothing effect*" is one of the important aspects of the Baldwin effect; as learning (or plasticity) makes the spiky fitness landscape much more like a standard hill. This is part of the reason that the Baldwin effect is described as learning synergistically aiding evolution; not only learning ontogenetically finds the summit, but it subsequently paves a way for evolution to find the same summit phylogenetically.

Smoothing effect is also known as the Baldwin "*expediting effect*"; as the result of the smoothing process, evolutionary search can look for the spiky summit in a similar manner to a normal Fujiyama landscape. This consequently speeds up the evolutionary process compared to the case without learning.

2.4 The Canalizing Effect

Costs of learning also bring another important aspect to the Baldwin effect. Basically this is what is described in Stage 3 of Simpson's description of the Baldwin

effect. Because of the smoothing effect, the population now evolves to reduce the costs of learning by increasing ineluctability of the trait; over the course of evolution, learnt (or plastic) traits are 'canalized' by the more ineluctable traits. This reduction of costs of learning is evolutionary more favorable, since such costs negatively affect one's fitness. Although this is often equated to Waddington's (1975) concept of "genetic assimilation", some clarifications should be required. This point will be discussed later.

Such an increasing process does not necessarily continue until the learnt behavior becomes completely ineluctable (thus no learning is required). If the costs of learning are sufficiently reduced compared to other factors, the selective pressure working on this process would be weak enough to suppress it; if there is no cost for learning, this type of process will not take place at all. Also, if the environment is ever changing so that the position of the summit on the landscape is shifting, such a process would not take place or greatly reduced.

2.4.1 Canalization

It is well known that wild-type organisms are, compared to mutants, generally consistent in their end-state phenotypes, while environmental conditions in which they are reared vary. This is an interesting property of the wild-type organisms, as mutants are more likely to be sensitive to epigenetic conditions. Waddington conceived this as an important property of organisms and termed it "*canalization*". In one of his earliest expressions of this idea, Waddington stated it as follows:

> [Developmental reactions] are adjusted so as to bring about one definite end-result regardless of minor variations in conditions during the course of the reaction.
> (Waddington 1975, p. 17)

He considered that this buffering effect against developmental perturbations must be quite common as wild-type organisms are amazingly constant even under various epigenetic environments. Empirically this has been attested. Wagner *et al.* (1997) provide a nice summary of the core facts on which the idea of canalization stands.

1. Mutations with a major effect on a quantitative character. . . increase the variance compared to the wild type.
2. A similar effect is observed with environmental disturbances.
3. Canalization is causally inhomogenous. Canalization of the effects of one gene does not imply canalization to mutations at another locus.

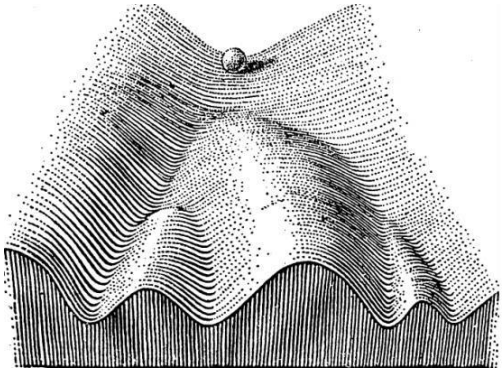

Figure 2.1: Epigenetic Landscape

4. The sensitivity of a character to genetic and environmental perturbations is correlated with its influence on fitness.

The first statement implies that mutations disclose genetic variations which have failed to be expressed in the wild type. This provides an important foundation for the assumption that wild-type phenotypes are generally canalized. This means that under wild-type phenotypes, some degree of genetic variation exists and it is higher than that of laboratory-reared mutants. A similar thing can be said about environmental perturbations based on the fact that wild-type phenotypes are consistent even though different environmental conditions are common in nature. The third statement provides some caution regarding the notion of canalization. Canalization is a notion which only captures a specific character of phenotypes. For example, the effect of canalization against heat shocks does not imply that the same type of canalization can be observed regarding different stresses (*e.g.*, salinity in given organisms). Finally, the most important point is that such buffering effects may contribute to fitness both positively and negatively; canalization stabilizes development, while it also means reduction of phenotypic variations on which selection works.

Often the general notion of canalization is expressed in Waddington's epigenetic landscape (Figure 2.1). The number of valleys on the landscape represents the range of reactions. Thus if it has a single, deep valley, the number of possible responses will be just one; an extremely robust development resistant to genetic/environmental perturbation is attained. On the other hand, if the landscape is flat, the developmental system has sheets of sensitive reactions; high plasticity exists. Canalization is the process of deepening these valleys.

In these cases, crossing the canals is a passive, and deterministic phenomenon. The genetic/environmental perturbation is the agent that changes the development, and the development is also pre-determined by the environment or the genetic background; there is a fixed set of reactions to such conditions. In other words, genetic/environmental conditions *induce* certain developments (and phenotypic end-states).

It is important to note that canalization is a type of dispositional concept; it describes the property of development in which different levels of causes may be involved. Waddington admitted:

> The notion of canalization is, therefore, intended to be a very general summing-up of a large number of well-known facts in genetics and embryology, all of which are summarized in the statement that the development of any particular phenotypic character is to some extent modifiable, and to some extent resistant to modification, by changes either in the genotype or in the environment.
> (Waddington 1975, p. 72)

2.4.2 Genetic Canalization

As seen in the above quotation, canalization is applicable to different levels of development. As canalization is confinement of phenotypic variations, classifying canalization based on the causes of such variations provides a good insight.

There are two sources of perturbations; genetic and environmental perturbations. Thus two different types of canalization can be defined, namely "*genetic*" canalization and "*environmental*" canalization. Within the context of insensitivity against perturbations, the two types of canalization are obviously similar to each other. However, they have different properties in many aspects. Here, we present concise summaries of these two types of canalization. First, let us look at genetic canalization.

Genetic canalization describes resitibility to genic differences. Topologically speaking, it is often the case that genetic variation does not perfectly match the corresponding phenotypic variation. In other words, phenotypic variations are sometimes smaller than genetic variations from which the former are derived. This is statically observable. On the other hand, genetic canalization in a more dynamic form is describable as a resistive property against mutational perturbations; when a phenotypic character is insensitive to mutational changes, it is due to genetic canalization.

Genetic canalization is triggered by two main causes. One is genetic neutrality, and the other is polygenic interactions. Kimura has argued that the majority of

mutations do not provide any impact on the phenotypic level (*e.g.*, Kimura 1983). For instance, two different alleles of a single gene express the same phenotypic trait. If such alleles are neighbors of each other (*i.e.*, the difference of the two genotypes is attributed to one mutation), then a mutation on the locus does not affect anything (*see* Appendix for a more detailed description). Since mutations are one of the sources of genetic perturbations, in this case the perturbation can be thought to be 'absorbed' by the neutrality. Thus neutrality is a mechanism of genetic canalization. If two or more genotypes are equally adaptive in terms of a specific trait and they are connected by mutations on a given locus (*i.e.*, all such 'neutral' genotypes are accessible by other neutral genotypes without passing any non-neutral genotypes), such a 'network' absorbs many mutations without expressing noticeable differences on the phenotypic level.

Another source of genetic canalization is polygenic interactions. Usually an expression of a phenotypic trait is not attributed to a single gene. Rather, interactions with other genes and environmental factors are common. In other words, it is unusual that a gene attributed to a phenotypic trait acts on its own without any genetic and/or environmental 'background' (*see* Section 5.2 for a detailed description of polygenic interactions).

This genetic background non-trivially determines the contribution of the gene. It is highly possible that, because of the background, replacements of alleles (due to mutations) do not affect the phenotypic level; the background works as a constraint on expressing the difference of the two. One example is a polyploid system. Many higher organisms have diploid inheritance systems. A diploid (or polyploid) system is in some sense redundant as it includes two competitive alleles in one locus; indeed the genetic value of a heterozygote is not exactly the average of the genetic value of the two homozygotes. This is somewhat mysterious, since those alleles are perfectly capable of containing genetic information on their own. In other words, fundamentally, each such allele in a locus is no different from an allele in a haploid system. This masking effect is called dominance; one of the alleles in a locus suppresses the other allele's expression. The dominated allele is called recessive.

By this mechanism, the dominant allele exclusively expresses its phenotypic trait. Thus, a diploid system attains the same result as a haploid system does. However, in contrast to a haploid system, because of dominance, mutations on recessive alleles generally do not provide effects on the phenotypic level. Hence a diploid system enables organisms to have some buffering ability against mutations (but not on dominant alleles), and the organisms are genetically canalized. This buffering effect is well-represented in inheritance of features such as albinism. In mammals, expression of the albino phenotype is due to a recessive allele on a specific

locus. Consequently, most hosts of the allele do not express this character simply because it is overruled by a dominant allele. Only when the dominance is broken by a homozygous combination, does it appear on the phenotypic level. If such genetic variation occurs in haploids, albinism will be more frequent. Thus, on a large scale, evolution from haploid systems to polyploid systems can be thought to be an evolutionary development of genetic canalization.

The other example of polygenic interactions producing genetic canalization is *epistasis*. Although we will delay detailed discussion regarding epistasis itself until Chapter 6, it is worth providing a brief explanation of epistasis and genetic canalization. In contrast to the system of polyploidy, epistasis describes non-linear polygenic effects of alleles on two or more loci; a specific phenotypic trait is affected by several genes in different loci. This type of polygenic interaction complicates the contribution of a specific gene to a specific phenotypic trait. If one phenotypic trait is determined by two or more genes, but the effects of each allele are additive, it is not called epistasis. Phenotypic individual differences are predictable from the additive effects of allelic substitutions.

Therefore in the case of an additive effect, variances may be described by examining a single locus. This effect is closely related to genetic canalization. For instance, if an allele that has a strong additive effect is selected, the genotype is, in practice, genetically canalized as other mutations' effects are weakened by the strong additive effect. In this case, other genes in different loci act as a background. On the other hand, if interactions of multiple genes are non-linear (*i.e.*, the effect is not additive), allelic differences may be completely canceled. For example, if other genes' influences dominate the phenotypic expression, allelic differences on a particular gene may not be reflected in the phenotypic trait. This produces a similar effect to genetic neutrality and consequently, the strong type of genetic canalization emerges.

2.4.3 Environmental Canalization

The term "environmental canalization" designates consistency of epigenetic development under different environmental conditions. Effects of environmental perturbations are determined by the following two factors. First, the intensity of the perturbations. Obviously, slightly different environmental conditions may not affect phenotypic variations (although the degree of intensity is a relative concept). If environmental factors become extreme, 'abnormal' phenotypic traits are often observed.

The second factor is plasticity. If a phenotypic trait is equipped with a range of reactions against environmental perturbations (*i.e.*, plasticity), environmental

canalization works to reduce the range of such reactions. Thus what is targeted here is quantitative phenotypic variation. As a result, while an organism can be plastic, it also can reveal some degree of insensitivity against environmental perturbations. In other words, while plasticity itself provides phenotypic variations, it also prevents oversensitive reactions to environmental differences.

Although this sounds somewhat contradictory, a metaphorical example provides a good insight; when a large building is about to be built on volcanic land, it has to be resistant to earthquakes. To attain this demand, there are roughly two possible solutions. One is making the building and its basics as hard as possible. The other solution is making the building somewhat flexible. By doing this, the building itself acts as a buffer and consequently absorbs quakes.

A similar argument is applicable to the case of higher-order cognitive abilities; while linguistic environment is largely different from person to person, the end-states of human linguistic abilities are surprisingly similar. Of course, measuring linguistic ability is hard and there are individual differences to some degree. However, comparing the degree of differences in one's growing environment with the degree of differences in one's linguistic abilities, naturally the existence of strong constraints on the range of adult competence must be admitted. In this regard, linguistic ability is thought to be environmentally canalized.

On the other hand, human beings are able to acquire strikingly different languages. What connects these two properties is plasticity of language learnability. Because language learnability is plastic, people can communicate with each other, even their personal experiences are idiosyncratic and could be enormously diverse (as long as they are reared in the same linguistic community). And because language learnability is plastic, one can learn any language in the world as long as one spends his childhood in a given linguistic environment[2] As in this example, cognitive abilities are often plastic and consequently strongly canalized.

2.4.4 Canalization and Ineluctability of Development

Theoretically, the two types of canalization have different impacts on evolution. As genetic canalization confines varieties of phenotype, it directly affects natural selection. Generally, strong genetic canalization is believed to impede the pace

[2]It is important to note that this argument does not hinge on the nature–nurture debate. First of all, any theoretical enterprises have to admit language acquisition is plastic. This is simply an undeniable fact. Secondly the environment canalization in language acquisition argued above is possibly attained by direct genetic influence, namely universal grammar (**UG**, which is assumed by generative linguists), *and/or* by a constellation of other cognitive abilities which, as a whole, support language acquisition (which is believed by functionalists and others). The important point is that environmental canalization in language acquisition is simply a description of the fact of language acquisition and itself does not provide any mechanism. This is also true for any type of canalization in general.

of evolution as it reduces phenotypic variations on which selection acts. Interestingly, genetic canalization itself does not necessarily reduce *genetic* varieties. As long as variances on the genetic level are not reflected on a phenotypic level, genetic canalization does not affect genetic variances. Thus in some cases, somewhat counterintuitively, genetic variation increases, while phenotypic variations are kept intact.

The example of polyploidy nicely describes this point; effects of mutations on recessive alleles are not ferreted out until a homozygous combination emerges. Therefore, in the case of albinism, for instance, the mutated allele responsible for lack of pigment is likely to be passed on to later generations as it does not affect fitness unless the same recessive alleles occupy the same locus. The recessive allele is autosomal; if one of the parents has the recessive gene and the other does not (or, obviously in the case that neither has the allele), none of their offspring expresses albinism. If both parents have the allele but it is dominated by the other allele, only one in four of their offspring will express it. And only in the case that both parents are albino, their offspring obligatorily expresses it. In most cases the recessive allele acts as a neutral gene, and consequently it may well go through the sieve of natural selection. This, of course, depends on some balancing effect, still the buffering effect of diploidy keeps or increases genetic varieties in the genepool. As epistasis is a more general case of polygenic interactions (and it possibly connects different alleles in different loci), both the buffering ability and neutrality are presumably much higher than in polyploidy.

Environmental canalization does not have a direct impact on selection. Its effects on evolution are more indirect. However, this type of canalization contributes to stabilizing development. Environmental factors during one's developmental stage could be enormously different on an individual basis. Therefore, buffering environmental perturbations (and consequently, stabilizing development) likely become an important property, as Waddington speculates. As in the case of genetic canalization, consistency in epigenetic development shows that environmental canalization is potentially a common property of organisms.

The other important aspect of environmental canalization is its relationship to plasticity. As argued above, plasticity is an important factor of environmental canalization. However, what Waddington conceived was a more primitive and fundamental relationship between plasticity and environmental canalization. He argued that the concept of environmental canalization essentially incorporates the notion of plasticity. For instance; "*The idea of canalization involves no more than that the course of development exhibits, in some way, a balance between flexibility and inflexibility*" (Waddington 1975, p. 81; Original italics).

The work of Schmalhausen (1949) nicely made a point of Waddington's intention. Schmalhausen argued that genetic systems do not perfectly determine phenotypes. At most, they set up a range of reactions. Within this range of reactions, phenotypic traits develop. Therefore, 'acquired' characters must also sit somewhere in this range. This range of reaction is known as the reaction norm (a detailed discussion will be given in Section 5.2).

As evolution goes on, the norm of reaction changes so that the range of reaction becomes narrower. Therefore, evolution primarily changes the reaction norm itself through stabilizing development. This is called the "*stabilizing effect*". Then the second aspect of the Baldwin effect, namely increasing ineluctability of a learnt trait can be considered as result of a narrowing the range of such reaction norm; through environmental canalization, a learnt trait can increase its ineluctability. This is the "*canalizing effect*". Importantly, in his introduction of the Baldwin effect, Simpson (1953) admitted that Schmalhausen's study is possibly the closest example of the Baldwin effect in the language of the Modern Synthesis.

2.5 Genetic Assimilation

2.5.1 The Basic Formulation

Waddington considered that plasticity in epigenetic development is normally regulated by a pre-existing environment canalization at an arbitrary point in evolution. When environmental conditions change, the pre-existing canalization is broken and the previously concealed plasticity (*i.e.*, variations) is revealed. Waddington considered that if the new environment is sustained for long enough, a new canalization process will set a new fixation point in the epigenetic environment. Consequently, plasticity is once again confined but at a different point. He termed this shifting process "*genetic assimilation*".

Waddington expresses the concept of genetic assimilation in the following manner: Suppose that an individual would express the phenotype P from the genotype G in the environment E, while if reared in environment E', G expresses the phenotype P'. Those features in which P' differs from P are considered 'acquired characters' and the resemblance between P and P' is considered 'inherited characters'. Then, suppose that, in environment E, individuals in population A typically exhibit phenotype P. Suppose also when we put a sub-population A' in environment E', individuals show phenotypes P'. Then $P - P'$ is an acquired character. Now, A' has continued in E' for a considerable amount of generations, let it return to E. In this environment, suppose that an individual in A' exhibits phenotype P''. Then the degree to which P'' resembles P' is a measure of the degree of genetic assimilation of acquired character $P - P'$.

With Figure 2.1 (p. 22), this formulation can be described as follows: Basically it changes the landscape so that the positions of the canals shift; initially an environmental perturbation pushes the ball out from the default canal. If this persists for a certain number of generations, eventually the landscape is deformed and the alternative canal becomes the default. Thus, when A was returned to E' from E, phenotypes (*i.e.*, P) will still resemble P'. Thus the notion of canalization is logically entailed by genetic assimilation; while the notion of genetic assimilation includes environmental changes, canalization simply denotes increasing of ineluctability.

2.5.2 Phenocopy

The above formulation is hardly likely to be testable under normal conditions. Waddington investigated it in the laboratory with various subjects with both artificial and natural selection. The most common example of his studies is expressions of *Drosophila melanogaster*'s crossvein.

If embryos of *Drosophila* are exposed to ether at a certain development stage, some of the embryos develop a type of phenotype that has wing-like structures (Figure 2.2, p. 30; from Waddington (1975)). Interestingly, while this type of *Drosophila* is obtained exogenously, a similar abnormality can be indigenously expressed due to a mutated gene, called the *Bithorax* mutant. This environmentally produced phenotype which mimics a genetically produced phenotype is called a "*phenocopy*" (Goldshmidt 1938). Genetic assimilation is based on the existence of such a phenocopy.

If over the generations, artificial selection is conducted so that only such phenocopies (in other words, ones that possess the acquired character) are prolific under the condition of ether vapor exposure, after a comparatively small number of generations, even without ether, the character, this time indigenously, is obtained (Figure 2.3, p. 30). In the graph, four lines are plotted; the x-axis shows the number of generations, and the y-axis designates the percentage of the bithorax reared without ether (some of the offspring are reared under a normal condition to check how many offspring innately expresses bithorax. Other offspring are reared under the stressed environment). Two types of experiments were conducted, one with selection *for* the bithorax-like phenocopy and the other with selection *against* such a phenocopy. As can be seen in the graph, around the 15th generation, in both types of selection, the proportion of indigenous individuals reaches almost 100% (or 0%).

Similar results were obtained by natural selection when larvae were fed with salt added to food in quantities sufficient to cause considerable mortality –a characteristic physical trait of salt-resistant flies started to be inherited within a small number of generations (this means that the number of salt-resistant flies was increasing by

Figure 2.2: a *Bithorax*-like Fly

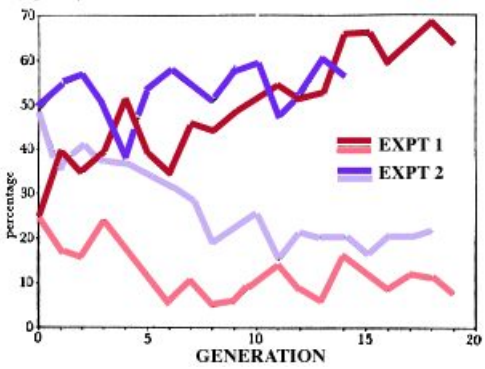

Figure 2.3: Evolution of *Bithorax*-like Fly

generations). These experiments show that the acquired characters are genetically replaced within a considerably short span; genetic assimilation.

Note that the existence of a phenocopy is a crucial assumption of genetic assimilation; if one (or nature) cannot find any phenocopy at all, he cannot even start selecting right individuals.

2.5.3 Genetic Assimilation & The Baldwin Effect

As one of the first developmental geneticists, Waddington attempted to integrate the evolution of acquired and inherited characters. This is basically in the same spirit as the Baldwin effect; both capture the flow of an evolutionary process from an environmental change to increasing ineluctability of development. Indeed, genetic assimilation is conceptually very similar to the Baldwin effect.

However, at the same time, it has been equated with the canalizing effect too. This is a type of common confusion rooted to genetic determinism; a process which increases ineluctability is instantly equated with increase in the contribution of genes. From this point of view, it is not difficult to consider development of a particular phenotypic trait as 'genetically assimilated' if its ineluctability is somehow increased.

Moreover, it should not be forgotten that because Baldwin himself provided no genetic-based explanation, researchers have implicitly assigned yet another meaning to the term. Indeed, Waddington himself criticized a lack of genetic basis in Baldwin's concept.

> The process... differs from the notion of genetic assimilation primarily because it considers the initial adaptation to the new environment to *be a nongenetic phenomenon on which selection has no effect.*
> (Waddington 1975, p. 89: emphasis added)

> This theory seems to be an impossible one. The acquirement of an adaptive modification in response to an environmental stress cannot, according to all our basic ideas of genetics, be due simply to a plasticity of the phenotype to which the genotype is quite irrelevant.
> (Waddington 1975, p. 89)

The ambiguity of genetic basis of the Baldwin effect, as a result, leaves a certain degree of freedom for interpreting Waddington's genetic assimilation not only as the canalization process itself but also as the 'mechanism' of the process. Consequently, a large body of recent studies have referred to Waddington's genetic assimilation as the *genetic mechanism* of the Baldwin effect. This tendency is fortified further by the work of Hinton & Nowlan (1987). As in normal GA simulations, only minimal components are required for what they are intended to reveal (*i.e.*, interactions between learning and evolutionary search). Thus in the simulation, what individuals inherit at the end of a run is exactly the same as what their ancestors have learnt. As the process is really a replacement of learning by genetically pre-specified traits, nothing but the term genetic assimilation is suitable for describing the process. However, as we will see in Chapter 4, such a mechanism is just one of the possible mechanisms the Baldwin effect may take.

In summary, the term is associated with at least three different aspects of the Baldwin effect –the Baldwin effect itself, the canalization process, and its mechanism. Together with Simpson's formulation of the Baldwin effect, Waddington's concept of genetic assimilation greatly enhanced scientific plausibility of the Baldwin effect. However, as apparent from the above discussion, there are good reasons to avoid use of this term in the context of the Baldwin effect. First, because of the word 'genetic' in the term, by referring to the canalization process, it is almost unavoidable to associate some process of '*being genetically hard-coded*'. However, as described in Chapter 1, such a genetically deterministic view is unnecessary for the

canalizing effect. Secondly, mixing the term in the mechanical sense with the phenomenal sense turns researchers' attention away from the fact that the mechanism of the canalization process in the Baldwin effect is pluralistic.

2.6 Lamarckian Inheritance and the Baldwin Effect

Because the canalizing aspect of the Baldwin effect states that learnt behavior is eventually taken over by a more innately predisposed trait, the Baldwin effect is often confused with Lamarckian inheritance. It is worth comparing the Baldwin effect with Lamarckian inheritance in this sense. Lamarckism is often expressed as a direct inheritance of characteristics acquired by individuals during their lifetime. In other words, there is a direct feedback channel from learning to genes through which information gained by learning can be reflected on genotypes in the next generation.

On the other hand, in the Baldwin effect, there is no such direct feedback channel. Learning is independent from evolutionary search and no information flow is allowed from learning to genes. While in Lamarckian inheritance, the feedback channel makes the information flow as circumfluent, in the Baldwin effect, the canalization process triggered by the costs of learning superficially connects learning and evolutionary search.

It has been attested in computer simulations that, in a highly fluctuating environment, these two behave quite differently. In Lamarckism, individuals can adjust their phenotype to increase their fitness immediately after an environmental shift. Although this inheritance mechanism attains both the canalizing effect and the expediting effect, if individuals adapt themselves to the current situation too greedily, they behave in an *ad hoc* manners and turn out to perform quite poorly in the overall dynamic environment. On the other hand, with the Baldwin effect, since learning and biological evolution are clearly separated, individuals can cope with the detailed changes at an individual level of learning, while to some extent keeping the generality. In this sense, the Baldwin effect can find a hillclimbing path even in a dynamic environment. As shown in below, Avital & Jablonka (2000) have developed this idea and termed "*the categorizing effect*". Sasaki & Tokoro's works make these points clear experimentally (*e.g.*, Sasaki & Tokoro 1997). Their works will be presented in 2.8.4.

However, the current majority of studies of the Baldwin effect also require a particular relationship between genotype and phenotype if the canalization process is to be induced. This condition, called "genotype-phenotype correlation" (**G-P** correlation), states that the search by learning has to be positively correlated to evolutionary search. In the scheme of fitness landscape, it is described that the

shape of a landscape on learning has to be highly similar (or the same) to that of evolutionary search. Although this condition is theoretically perfectly conceivable, it requires a somewhat unpractical assumption in the relationship between genotype and phenotype especially in higher-order traits such as behavioral abilities. As this is the major motivation to reconsider the Baldwin effect and leads us to consider a new mechanism of the Baldwin effect which is proposed in this thesis later, we will leave this topic here.

2.7 Advances in the Baldwin Effect

In *Animal Tradition* (2000), biologists Eytan Avital & Eva Jablonka spend a whole chapter on the study of possible impacts of learning on evolution. Naturally, the Baldwin effect is in the scope of the topic. They propose three significant impacts of the Baldwin effect. In this section, these concepts are briefly discussed.

2.7.1 The Categorizing Effect

As discussed already, increase of ineluctability of a phenotypic trait through the canalization process rarely becomes extreme. In other words, the canalization process is terminated so that still some degree of learning contributes to acquiring a given phenotypic trait. Although some reasons are conceivable, one plausible reason is that environmental conditions fluctuate somewhat so that no definite fixation point of phenotypic value of the trait is available. Therefore, for organisms to be adaptive in different conditions, it is better to preserve some plasticity for the trait. If the range of environmental shift is large, the canalization process will scarcely proceed; by keeping a large degree of freedom in epigenetic development (*i.e.*, plasticity), organisms can cope with a variety of environmental conditions. On the other hand, if environmental shifts are confined within a certain range, the canalization process may proceed where the innate predisposition can cover the 'largest common denominator' (**LCD**) of such environmental shifts.

Consider the relationship between this 'LCD' and actual environmental conditions. Each environmental condition is an individual instance while the LCD is, as the name designates, the overall property which such conditions share. In other words, the LCD 'categorizes' such environmental conditions against other possible conditions. This process is experimentally attested by Sasaki & Tokoro reviewed in Section 2.8.4.

An interesting point is that when this process takes place in the domains of behavioral, or psychological/cognitive behavior, organisms are 'unconsciously' equipped with such a mental categorization. Dor & Jablonka apply this concept in their theory of language evolution and speculate that the Baldwin effect can contribute to

create syntactic categories evolutionarily. Their studies are reviewed in Section 3.1.4.

2.7.2 The Assimilate-Stretch Principle

The second concept concerns a meta-application of the Baldwin effect. Avital & Jablonka conceive that a behavioral sequence can be lengthened without altering learning ability. To see how this works, imagine an organism capable of learning a sequence of four consecutive behavioral actions. Suppose also that the learning capacity of the species necessary for this sequence is constrained so that the learning ability itself would not evolve. Imagine this as a case of a bird and the target behavior is nest-building; to build a nest, the birds have to learn, say four behavioral steps. If there is a consistent pressure for efficient and reliable nest building, it is conceivable that the canalizing process takes place. Through the process ineluctability of one of the sequential steps becomes high so no more learning has to take place in this step.

Avital & Jablonka consider that under this condition, it is conceivable that organisms are now capable of adding an additional adaptive learnt action to the remaining three, since the learning capacity is 'freed' from the first step. Then, the population is now equipped with five consecutive actions. Of course, there is no reason to confine this process to occurring just once; it is completely conceivable that this type of process takes place cyclically. They term this process "assimilate-stretch".

This is a powerful tool of the Baldwin effect. With this principle, now one can consider not only increase of ineluctability, but also evolution of behavior which is not directed by genetic evolution but learning; it continuously guides the evolutionary process so that as a result a highly sophisticated behavioral trait is attained.

In this principle, there is at least two significant prerequisites are implicitly assumed. First, the species has to be equipped with some non-genetic capacity to create new behavior. If new behavior is genetically invented (*e.g.*, through mutations), there is no need for learning to take place in first place. Secondly the learning capacity is fundamentally domain-general; the capacity has to be applicable to different types of behavioral actions.

As we will see in Chapter 4, these prerequisites are hardly met under the current model of the Baldwin effect which Avital & Jablonka also adopt; for the model to work, learning has to be genetically bounded by the same genes which regulate the innately predisposed phenotypic trait. This simply contradicts the above assumptions. In Chapter 6, we propose a new mechanism which does not require this

regulation. Consequently, under the new model, this assimilate-stretch principle plays a vital role.

2.7.3 Switching Modalities

Avital & Jablonka also consider that learning causes the canalization process by exapting previously unused regions in the brain. They consider that a newly learnt behavior can reactivate once dysfunctional cognitive capacity. This is a case of canalization since such dysfunctioned capacity is more genetically hard-encoded than the learnt behavior, they consider. Avital & Jablonka assume that the virtually degenerated visual capacity of the Palestine mole rat has been 'reused' by the enhanced auditory system through the Baldwin effect; it is known that the brain regions that seeing mammals use for processing visual information is extensively used for auditory information in Palestine rats. They suspect that this switching process was initiated by a change in habits. It is highly unlikely that mutational changes made the rats blind and that then they change their habits. When their ancestors chose to live underground, the physiological capacity of vision was not dysfunctional. Over the generations, the function was lost. However, the visual processing capacity was kept and eventually the enhanced auditory system 'parasited' on it.

Importantly, Avital & Jablonka seem to identify this switching process as increase of genetic attributions (they explicitly use the term "genetic assimilation" in the sense of the canalization process). Because of this, their formulation of this process suffers from the same difficulty described in the assimilate-stretch principle. However, given the discussion given in Chapter 1, this does not have to be the case; as long as newly incorporated modalities increase ineluctability, such modalities have to be no more genetically attributed than previous ones. In the same way as the assimilate-stretch principle, the new mechanism can circumvent the problem and makes this exaptational aspect of the canalizing effect highly important.

2.8 Computational Studies

2.8.1 Evolution and Genetic Algorithms

Before introducing computer simulations regarding the Baldwin effect, first we briefly discuss GAs in general[3].

GAs are a type of algorithm whose primary purpose is searching solutions to a problem. As is obvious from its name, GAs adopt the notion of Darwinian biological evolution; the combination of natural selection and natural genetics enables populations to adapt to environments. Given the fact that adapting to a natural

[3]This section is based on Shapiro (2001)

environment is an enormously complicated task, evolution has done an amazingly good job for billions of years; a wide variety of species adapt to the tremendous range of environmental conditions in a fairly short time. This adaptation to environments is, from the informatic point of view, considered as a type of problem solving. An evolutionary process is equivalent to a system where a group of agents is searching solutions to a problem. This inspired John H. Holland (1975) to implement the exquisite mechanism of nature *in silico*.

The fundamental principles of GAs are basically the same as in evolutionary theories in the Modern Synthesis; randomly produced variation on underlying structures is trimmed by selection according to their manifested entities. Subsequently, terms used for a GA are adopted from genetics. In a GA computer simulation, individuals (or agents) form a population and generations. Individuals' underlying forms are genotypes (or a chromosome) and their manifested entities are phenotypes. Genotypes are made of genes. A certain location of the chromosome is called a locus and possible genes appearing on a given locus are called alleles.

As described in Section 2.3, the search space is often described as a landscape. The best genotype (*i.e.*, the best solution) has the highest point on the landscape. Typically, neighborhood genotypes of the best genotype have also high fitness. This makes the fitness landscape Fujiyama.

The basic mechanism of the algorithm is as follows:

Representation A genetic algorithm itself does not specify how problems and solutions are encoded. Therefore, we have to determine the representations of both underlying structures and manifested entities in a GA. In a natural biological system, underlying structure is the genotype which is ultimately composed by DNA, and the manifested entity is the phenotype which can be a physical entity, behavior, or knowledge/information. In a GA, both the underlying structure and the manifested entity are ultimately expressed in binary but their intermediate representations (*i.e.*, where human beings actually encode a program) can be other symbols. Regarding the underlying structure in a computer simulation, the unit of underlying structure is usually equivalent to genes, but it can be at a more minute level, *e.g.*, DNA. It is often the case that the underlying structure does not represent diploidy or multiploidy. However, where it is necessary, such properties can be represented. In terms of the manifested entity, if the problems and solutions are mathematical or informatic, the manifested entity expressed

in binary is directly comparable to the solutions. On the other hand, in the case of, say, language evolution, manifested entities in a simulation are not directly comparable to real phenomena. Therefore, some interpretation of the representation is necessary.

Fitness Function The fitness function is the criterion on which selection works. In a GA, the function measures the *objective function* which determines the goodness of a given manifested entity. In problem solving, the object is the best solution. Usually, the fitness function is a monotonic function of the objective function. However, there are some types of GA in which the objective function is not immediately apparent. In the case of co-adaptation, for example, fitness is calculated based not on an absolute standard, but on a relative standard (*i.e.*, how well an agent behaves among others). This is because selection depends on frequencies of phenotypes in a population. Suppose that phenotypes are strategies of *RoShamBo* (the Rock-Paper-Scissors game) or Mediocre played among agents. In such a case, there is no best strategy; goodness of a particular strategy depends on the frequency of other strategies in a given state. Therefore, the objective function dynamically changes as the simulation progresses. Under this condition, the objective function cannot be overtly expressed, while the fitness function is explicit. As an important part of the utility of a language depends on the frequency of usage in the whole population, it is this type of evolution that we have to consider. We will come back to this point in Chapter 6.

Population Dynamics A GA requires an aggregation of agents. Just as in a natural system, agents are mortal, and some of the agents produce offspring. During the reproduction process, operations will take place so that this cyclic process produces variation on underlying structure. Genetic operators (*i.e.*, *mutation* and *recombination/crossover*) produce heritable variances, and *selection* culls in such a way that the agents in the next timestep become more adaptive. These operators are described below:

1. **Selection**: Based on the fitness function, adaptive values are assigned to agents. Then based on the values, some agents are selected and carried into the breeding process. There are various types of selection mechanisms.

2. **Recombination/Crossover**: This is an operation that recombines a pair of parents genes to produce offspring. Typically, the two genotypes are cut at the same locus and swapped. However, other methods also exist (*e.g.*, multi-point crossover)
3. **Mutation**: This operation randomly changes values of alleles in a genotype. Typically this is kept to a low probability.

In an ordinary GA, an individual can normally occupy only one location on a landscape in his lifetime. In other words, evolution allows individuals to search with only one chance; when the individual is reproduced, his search is virtually finished. Therefore, mutations and recombinations are responsible for the search. This is somewhat obvious as in a GA, producing a variation (*i.e.*, searching) is exclusively attributed to these operators. Mutations and recombinations (especially the latter) are potentially able to produce great variation within a few generations, and the evolutionary search based on those operations is often called "global" search.

In this context, plasticity/learning is considered as an increase of the phenotypic variation within a single generation; the individuals are able to search a landscape in their own lifetimes. Typically, learning allows the individuals to search close to their original phenotypes. Learning is called 'local' search as opposed to the 'global' search of evolution. The synergistic effect of the global and the local searches is the Baldwin effect. In the next section, the seminal work of Hinton & Nowlan is described.

2.8.2 Hinton and Nowlan

In this section, the most influential computational study regarding the Baldwin effect, namely the study of Hinton & Nowlan (1987) is addressed.

Hinton & Nowlan used a GA to demonstrate how the effect works. As noted in the above section, computer scientists often describe evolution in a GA as 'hill-climbing' and view its search space as a 'fitness landscape'. The global optimum is equal to the highest position on the landscape. Thus the higher the position, the more prolific the agent will be. If the fitness landscape is gradual (*i.e.*, a Fujiyama landscape), agents at higher levels can reproduce more offspring. The process of evolutionary search, then, looks like an Everest climb. When explorers climb Everest, they set a couple of base-camps before they reach the pinnacle; they gradually climb up the mountain from those base-camps. In evolutionary simulations, agents standing on a higher mountainside can reproduce more children. This is equal to the parent setting its own base-camp on the position. By reproduction, the parent deploys its children around the landscape since the children may have slightly different genotypes from their parent due to genetic operations. Some of them might

locate at a higher position. Then the particular child can set a new base-camp. The higher position the child occupies, the more likely the child will set its own base-camp. Eventually we can expect that an agent will reach the top. Since any agent who occupies the top is the most prolific, eventually the whole population converges at the top (= the global optimum).

What Hinton & Nowlan did in their fitness landscape was, however, to put a spiky summit on a large fitness landscape instead of a "Fujiyama" and flatten out all the other space. In other words, no matter how close to the spike a position is, any other positions on the landscape assign the lowest fitness value[4]. This is often called a "*needle-in-a-haystack search*" since in this circumstance, evolutionary search is no longer hillclimbing (Maynard Smith 1987). Since agents are looking for the spike blindly, their searches would not show a convergence toward the spike. This genetic search mechanism makes populations move globally, rather than gradually pining down to a specific location (*i.e.*, the spike) in the fitness landscape.

In their simulation, each agent is represented by a string of twenty characters. This is the agent's genotype. Each locus along the string takes $\boxed{0}$, $\boxed{1}$ or $\boxed{?}$. In one agent in the first population, roughly five $\boxed{0}$ alleles, five $\boxed{1}$ alleles and ten $\boxed{?}$ alleles are randomly assigned (*i.e.*, the frequency of the alleles are 0.25, 0.25, and 0.5 respectively). The proportion of these alleles changes over the simulation by the recombination mechanism described below. There is a certain period in each generation at which all $\boxed{?}$ alleles of all agents would be converted to either $\boxed{0}$ or $\boxed{1}$ randomly. This process is considered as a learning period. Each agent is assigned 1000 learning trials. This is roughly equal to the total number of possible combinations of $\boxed{0}$ and $\boxed{1}$ expressed by $\boxed{?}$ alleles in the initial population; there are roughly ten $\boxed{?}$ alleles in a genotype in the initial population, thus, by and large, 2^{10} (=1024) possible phenotypes can be derived from the genotype.

These two different modes of search implement both evolutionary search and learning; individual fitness search by modifying phenotypes (= learning) and population fitness search by modifying genotypes (= evolution). Note that phenotypes are never written back into the genotypes nor passed to the offspring (if so, it is Lamarckian). The recombination mechanism of the characters was introduced so that the population could evolve within their genetic pool. This is the production process of a new individual through the splicing together of genotypes from two 'parents'.

Since the global optimum is on the landscape grid of twenty $\boxed{1}$, if an agent succeeds in having all $\boxed{1}$ phenotypes, it is considered as a learning success. Fitness

[4]Indeed, only two fitness values exist on this mode –the high fitness value (the spike) and the low fitness value (any other positions).

value is assigned to the agent based on the number of trials –the less trials the agent has used, the higher the fitness value assigned. The lowest fitness value is assigned to those who fail to learn the global optimum. The fitness function is given as follows:

$$FITNESS = 1 + \frac{19n}{1000}$$

In this formula, n designates the number of learning trials remaining after the individual has successfully learnt the optimal setting. The innately fully specified individual (*i.e.*, all loci are occupied by $\boxed{1}$ alleles) is twenty times fitter than those who cannot learn the optimal setting at all.

Note that, this fitness function effectively smooths the spiky fitness landscape; in contrast to non-plastic populations, the plastic population smooths the fitness landscape by virtue of the learning search –the expediting effect. Recall that under these circumstances, for non-plastic agents, any positions except the global optimum are equally sterile. However, for those plastic agents, the closer they start to the optimum, the better chance the agents have of achieving higher fitness values. Consider, for example, three agents: One has three $\boxed{0}$ alleles, five $\boxed{1}$ alleles and twelve $\boxed{?}$ alleles, another has twelve $\boxed{1}$ alleles and eight $\boxed{?}$ alleles, and the other has eighteen $\boxed{1}$ alleles and two $\boxed{?}$ alleles. The first agent has no chance of reaching the global optimum –it has fixed $\boxed{0}$ alleles. The second might reach the global optimum but the chance is slim –one in 2^8 (=256). 1000 learning trials may be sufficient, but its fitness would be far from the optimum. However, the third agent has a high probability of reaching the optimum in early trials. Then he has a better chance to be prolific. Therefore, over the course of evolution, the population is encouraged to have less and less plastic alleles but more and more fixed alleles of $\boxed{1}$ (*i.e.*, the canalization process).

With this model, Hinton & Nowlan concisely show that both the expediting and the canalizing effect (*i.e.*, the Baldwin effect) indeed take place; this simulation will be replicated in Chapter 9. Note that, in the simulation, the search spaces of evolution and learning are the exactly the same. That is, when some of new offspring genetically get closer to the optimal setting, they are also closer to the setting on the learning space. This is a case of G-P correlation.

Soon after this study was published, the biologist John Maynard Smith positively reacted. He summarized the study and left the following comment:

> To use their (Hinton & Nowlan's) analogy, finding the optimal neural set in the absence of learning is like searching for a needle in a haystack. With learning, it is like searching for the needle when someone tells you

when you are getting close.
(Maynard Smith 1987, p. 762)

2.8.3 Follow-up Discussions on Hinton & Nowlan

On one hand, the small simulation of Hinton & Nowlan (1987) made a considerable impact on the study of learning and evolution not only in computer science but also in other areas such as evolutionary biology, cognitive science, robotics, and evolutionary linguistics. Their simulation is so simple that it allows researchers to apply other simulations to it.

On the other hand, however, there is a point in the result of this specific simulation that puzzles researchers; [?] alleles are not fully canalized by [1] alleles so that the perfect ineluctability is attained at end of the simulation. Rather the result of the simulation on the original article showed a very slight decrease of [?] alleles and no further dynamics are observed. This is a somewhat puzzling problem. From the perspective of the Baldwin effect, the expediting effect took place in the simulation; the first step of the Baldwin effect. Indeed, it is through this effect that the simulation made the most significant contribution. Compared to a non-learning population, the learning speeds adaptations of the learning population. On the contrary, after the population eliminates [0] alleles from the genepool, no further evolutionary dynamics take place. In other words, a slight decrease of [?] alleles is observed in early generations, and it quickly reaches a steady state. This means that almost no canalizing effect took place in the simulation, thought to be the second stage of the Baldwin effect.

This is somewhat counterintuitive as reducing [?] and increasing [1] alleles is certainly the best strategy the agents could take. Hinton & Nowlan explained it as due to weakening of selective pressure. Although indeed high fitness in later generations reduces selective pressures reducing [?] alleles, this explanation might not be the major reason; the graph showed almost half of the loci in a genotype would be occupied by [?] alleles. This means that on average, with a probability of only 2^{-10}, an agent can attain the all-1 configuration in its phenotype; it is certainly not the optimal genotype. Thus it is more likely that some other factor works in this blocking of the canalization process.

While in the article Hinton & Nowlan did not mention the exact architecture of the GA used in the simulation, given the totally flat lines starting around generation 20 on the graph, and no mention of the mutation rate, the architecture most likely does not have a mutation mechanism. Based on this observation, Arita (2000) replicates Hinton & Nowlan with various mutation rates. Basically he assumes that the reason for the stagnation is the lack of mutation. To investigate this

possibility, first he conducts a complete replication of Hinton & Nowlan. The results reveal that the case Hinton & Nowlan presented in the original article is rather rare; multiple runs of the simulation show that the relative frequency of [?] alleles in a genotype is on average 0.2, instead of 0.5 in later generations. Moreover the standard deviation of the runs shows slight increase with the lapse of time. This indicates that the variance of the frequencies of [?] alleles increases in later generations in different runs. From the results, Arita assumes that the result of Hinton & Nowlan is a sort of special case in which the genepool was filled by a single genotype, so recombination could not produce any further variation. Given this, he conducts a further simulation in which the mutation mechanism is added. With various mutation rates examined, Arita finds that when the mutation rate is roughly at 0.001, the frequency of [?] alleles could go down to 0.1. Besides the decrease in frequency, its standard deviation is also significantly small. This implicates that selection properly gets a grip of genetic variations caused by mutation; a typical co-operative work of selection and mutations. However, if the mutation rate is more or less than 0.001, mutation starts competing with selection. What selection culled in previous generations is returned to the genepool by mutations; a small amount of genetic drift begins.

Given the simulation with the 0.001 mutation rate, Arita also studied the variation of genotypes with the result. The genetic variation radically drops when [0] alleles are expelled. During this period, the number of genotypes once drops to less than 10 types. After this process is completed (*i.e.*, no [0] alleles remain in the genepool), variation increases up to 30 genotypes in the genepool. This process goes hand in hand with decreasing [?] alleles. From the result, Arita explains that the Baldwin canalizing effect gets its grip when the population is filled by plastic individuals.

Harvey (1993) approaches the same point (*i.e.*, the persistence of [?] alleles) from a different path. Using diffusion equations, he explains the phenomenon as the consequence of genetic drift.

In Chapter 9, yet another possibility is examined with a replicated simulation. That is, as Hinton & Nowlan originally assumed, the shallow curve of the canalizing effect is due to their selection mechanism. Although they did not specify what type of selection mechanism was utilized, replicative studies reveal that a far less frequency of [?] alleles is obtainable when a relatively strong selection mechanism is adopted. This will be reported in Chapter 9.

2.8.4 Sasaki & Tokoro

The Lamarckian inheritance mechanism has been discarded for more than a century. While still some researchers are attempting to provide scientific data supporting Lamarckian inheritance (*e.g.*, Steele 1998), most researchers show somewhat allergic reactions towards this type of argument.

As noted earlier, in this context, the Baldwin effect is easily misconstrued as a type of Lamarckian inheritance system. Both Lamarckian and Baldwinian models of evolution advocate the synergy of learning and evolution, and importantly, claim that learnt traits could become a part of organism's innate predisposition. Of course these two processes are completely different, yet the similarity of the causality (*i.e.*, the synergy of learning and evolution causes canalization) confused some researchers for a long time. This situation has been improved by the introduction of computer simulations in biology; abstractions of biological mechanisms in computer simulations saliently present the impact of the Baldwin effect without the Lamarckian inheritance system. More than fifteen years after Hinton & Nowlan's simulation, the difference of the mechanism adopted in the Baldwin effect from Lamarckian inheritance seems to be well understood.

However, it is not clear exactly how the two modes of inheritance are different in their impacts on evolutionary profiles. This state of affairs is somewhat understandable since Lamarckian inheritance is physically impossible at least in general; subsequently there is comparatively little motivation to investigate this unscientific mechanism. What is unwarranted is the fact that this is also true in computer science, even though it is virtually free from the dogmas in biology, and Lamarckism is easily implemented. Putting the question in a different way, it is interesting enough to ask ourselves why nature selects the Darwinian inheritance mechanism on which the Baldwin effect stands rather than the Lamarckian system; it might not be impossible for nature to create Lamarckian types of inheritance mechanisms. It could be merely due to an accidental factor, but given that the Lamarckian process intuitively has immediate advantages, the comparison between the Baldwinian and the Lamarckian processes should be investigated. In this section, we look at works of Sasaki & Tokoro (and Yamamoto in one article) that investigate the question raised here (Sasaki & Tokoro 1997, Sasaki & Tokoro 1998, Sasaki & Tokoro 1999, Yamamoto *et al.* 1999). All these three articles share the same type of simulations. In this section, the first article (Sasaki & Tokoro 1997) is mainly presented.

Sasaki & Tokoro (1997) design an agent model in which agents evolve based on a GA. Each agent consists of a chromosome, a neural network, an output module

called "*Action Decision Module*" (**ADM**), and 500 units of 'life energy'. The neural network has three layers; five input nodes[5], three intermediate nodes, and four output nodes. Every initial connective weight in the network is encoded in the chromosome. Thus there are 27 genes in the chromosome (fifteen genes encode the weights between the input and the intermediate nodes, and twelve genes encode the weights between the intermediate and the output nodes).

A world populated by 100 agents contains two types of materials (although it is not in the original paper, for convenience, henceforth we name the materials mushrooms); 'edible' and 'poisonous' mushrooms. Each mushroom is encoded by an array of six bits. The task of the agents is discriminating those two types of mushrooms and based on the discrimination, determining their action; either 'eat' or 'discard'. The discrimination task is processed by the neural network. An array from a mushroom is fed into the network as an input. However, decisions are not solely made on the discriminated information. The information is passed onto the ADM where final decisions are made. The content of the module is a type of Boltzmann function. Thus an actual action can contradict what the neural network produces. Subsequently the parameter which controls the temperature of the distribution acts as the degree of 'adventurousness' of the agent. In the simulations, it is kept at a reasonably low value but not zero. When an agent eats edible material, she will get 10 units of life energy, if she mistakenly eats a poisonous material, 10 units will be subtracted from it. If no action is taken, nothing happens.

Learning is also conducted based on results of agents' actions. A reinforcement learning framework is used with a combination of back-propagation learning. The process is iterated a sufficient number of times. However, with the 'no action' decision, neither learning nor addition/reduction of life energy takes place. The amount of life energy in an individual at the end of each generation serves as fitness. Based on the amount, selected agents are carried into the reproductive process.

To implement both Darwinian and Lamarckian inheritance systems, two types of reproduction process are designed on top of the mechanism described above. Darwinian inheritance is implemented as usual; select two individuals and pass them into the recombination and mutation operations to produce germlines. No feedback from learnt knowledge is reflected on the germlines. On the other hand, to implement the Lamarckian inheritance system, parents' genotypes are modified at the beginning of the reproduction process; information of the final connective weights is copied onto the genotypes. From the genotypes, germlines are produced

[5]This is most likely a mistake in their description since mushrooms are, as decribed below, encoded by an array of six bits. Consequently, the total number of gene in a given indivudal would be 31, instead of 27 as described below.

and passed into their offspring through the recombination and mutation processes. For one crossover operations, between zero and four cross points are randomly set at random positions. A mutation takes place with a 5% chance. When it happens, the value specified on a given locus is randomly changed between the range of ± 0.5.

By using the model, two types of experiments are conducted. The two experiments are different regarding how dynamically the environments change. In the first experiment, the discrimination rule between edible and poisonous mushrooms does not change, while their repertoire dynamically changes. More precisely, in the experiment, edible and poisonous mushrooms can be differentiated based on the first three bits, regardless of the rest of a given array (consider the rest of the part to be covered with the '*' *don't care* symbols). Among the 2^3 (=8) possible types of mushrooms, four are designated as edible, and the other four are designated as poisonous. However, in any one environment, two edible and two poisonous mushrooms appear. Such environments periodically change every 20 generations (Figure 2.4). Although 36 (${}_4C_2{}^2$) possible environments can be created, Sasaki & Tokoro use just six types environments (Figure 2.5). In all of these environments, edible and poisonous mushrooms are discriminable by only two bits; and the same value is shared by both edible and poisonous in the third bit (note that the ordinal numbers do not correspond to positions of loci on the array). However, agents 'know' neither the existence of the noise bits (*i.e.*, the '*' symbols) nor the location of the third bit in the first three loci. The result is shown in Figure 2.6 (p. 46) and Figure 2.7 (p. 47)[6].

The result is remarkable. The Lamarckian agents immediately adapt to every new environment because of the immediate feedback from the manifested structure to the underlying structure (*i.e.*, from learning to genes). However, this local adaptation does not lead the agents to a global adaptation; fitness of the Lamarckian agents endlessly oscillates and no directional adaptation occurs across the environments. The frequency of the oscillation corresponds to the frequency of the environment changes (*i.e.*, every twenty generations). Any evolution in a global sense has not occurred.

On the other hand, while up to 1000 generations of the Darwinian agents are less fit than the Lamarckian agents, they show fairly normal hill climbing; they succeed in steadily increasing their fitness. And around the 1000^{th} generation, the Darwinian agents outperform the Lamarckian. Although they too are oscillating to some extent, their range is far narrower than that of the Lamarckian, and it is eventually compressed. This indicates that the synergy of learning and evolution

[6]All figures in this section are reprints from Sasaki & Tokoro (1997).

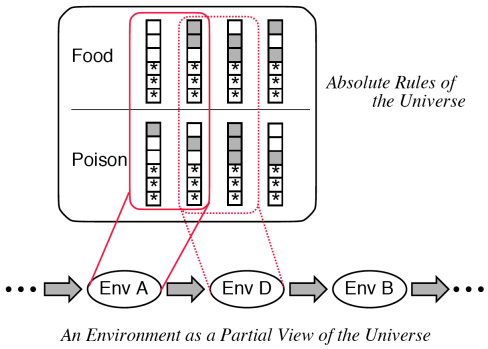

Figure 2.4: The Basic Model

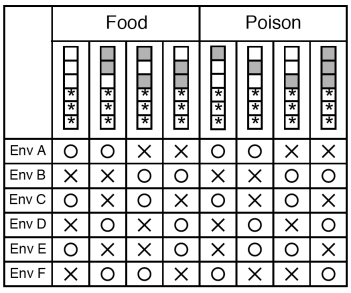

Figure 2.5: Experiment 1 –An environment where only partial information is available

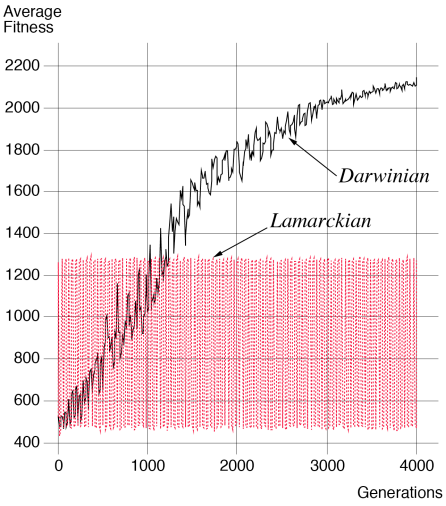

Figure 2.6: Experiment 1 –The average fitness of 0–4000 Generations

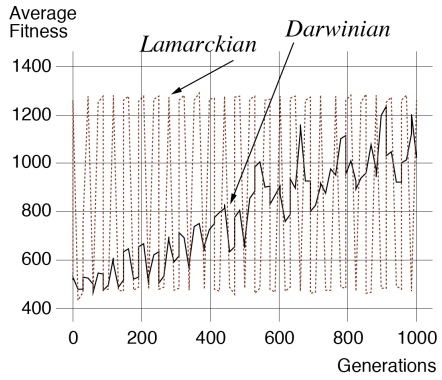

Figure 2.7: Experiment 1 –fitness The average fitness of 0–1000 Generations

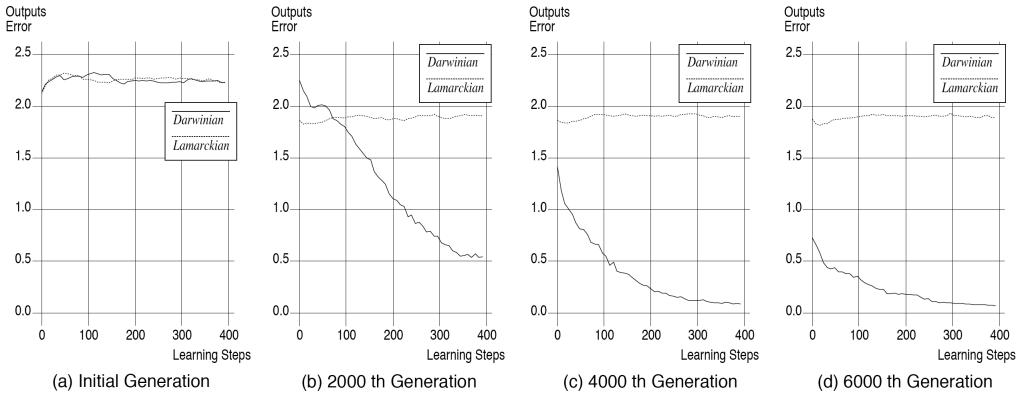

Figure 2.8: Experiment 1 –The changes in learning curves through generations

enables the Darwinian agents to find the 'hidden' categories of the mushrooms. Crucially, this evolutionary profile strongly indicates that the agents benefit from the Baldwin effect, especially from the canalizing effect[7].

To investigate this nativisation effect, a small simulation is conducted. From the first experiment, both types of agents at the first, 2000^{th}, 4000^{th}, and 6000^{th} generations are selected and trained in an environment with the complete set of mushrooms (*i.e.*, with the eight mushrooms). Figure 2.8 (p. 47) shows that in typical profiles of 400 learning steps while the Lamarckian agents in every generation keep the high error rate, the Darwinian agents decrease their error rate. This fact is most comfortably interpreted as the result of the Baldwin canalizing effect; Darwinian inheritance allows gradual canalization of learnt knowledge. Due to the rapid environmental changes, agents incorporate a fraction of the knowledge into genes at any one period; they fail to fully reflect what they learn about a given environment in their genes. However, the periodical appearances of environments enable them to gradually generalize the types of mushrooms. While the partial

[7]Curiously enough, Sasaki & Tokoro only lightly touch the Baldwin effect in (1999).

disclosure of the whole set of mushroom types tosses about the Lamarckian agents, its influence is somewhat masked by the 'failre' of canalization. It is this failure that enables the agents to adapt the entire world in a long run. If the duration of a period were long enough so that the agents could enjoy the canalizing effect enough within the period, this type of evolutionary profile would not be available.

The reason for the Lamarckian agents' failure is exactly found in this point. The experiment is explicitly designed for the discrimination task to be a two-bits parity problem (*i.e.*, the *XOR* problem). Although learning quickly finds the solution within a few generations, after this period what they learnt becomes not only useless but also even harmful; the knowledge is too specific for the previous environment and it cannot be applied to the new environment. Moreover, after moving into the new environment, Lamarckian inheritance quickly 'washes out' the previous knowledge. Then agents go back to the beginning of the circle; immediately adapting to the new environment and soon being deserted in yet another new environment. This also means that the inheritance system allows no global adaptation. This is a rather ironic fact when we consider what Lamarck intended to describe.

Sasaki & Tokoro conduct a further investigation regarding the influence of dynamism of environment on the Darwinian and Lamarckian processes. In the second experiment, mushrooms are differentiable by only comparing the first two bits out of five. Therefore there are four possible types of mushrooms: two are edible, the other two are poisonous. This time all the four types of mushrooms are introduced in each environment. After every 50 generations, however, the codes of the edible and the poisonous mushrooms are completely swapped. That is, after a given environment, the edible mushrooms become poisonous, while the poisonous mushrooms become edible; the discrimination rule itself changes (Figure 2.9, p. 49). The result is shown in Figure 2.10 (p. 49) and Figure 2.11 (p. 49).

The result is similar to the first experiment; while Lamarckian agents fail to find the globally optimal solution, Darwinian agents are certainly on the course of it. However, there are also some differences between the two experiments. First, the average fitness of the Lamarckian agents in the second experiment is far worse than the first one. The reason for this is probably found in the nature of neural network architecture in general. The discrimination task is basically *XOR*. When the rule is swapped after 50 generations, agents have to adjust connective values. However, as the Lamarckians strongly carry out their adjustment of connective weights for the previous environment, the degree of adjustment required for the new environment becomes so radical, the agents cannot cope with the situation. Moreover, while struggling to find the answer, yet another new environment comes into effect.

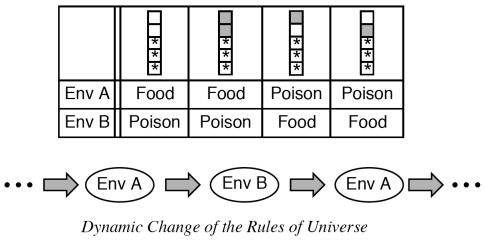

Figure 2.9: A dynamic environment where the rule changes

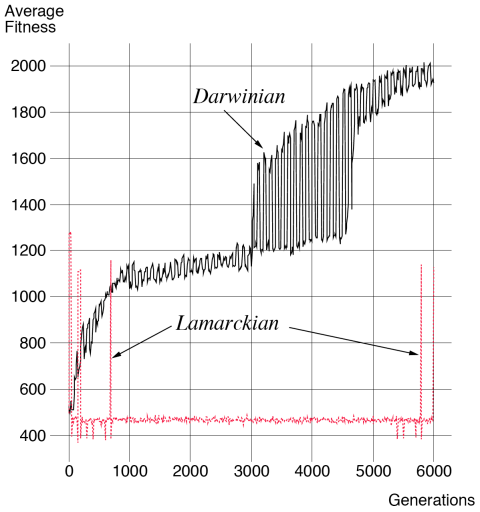

Figure 2.10: Experiment 2 –The average fitness of 0–6000 Generations

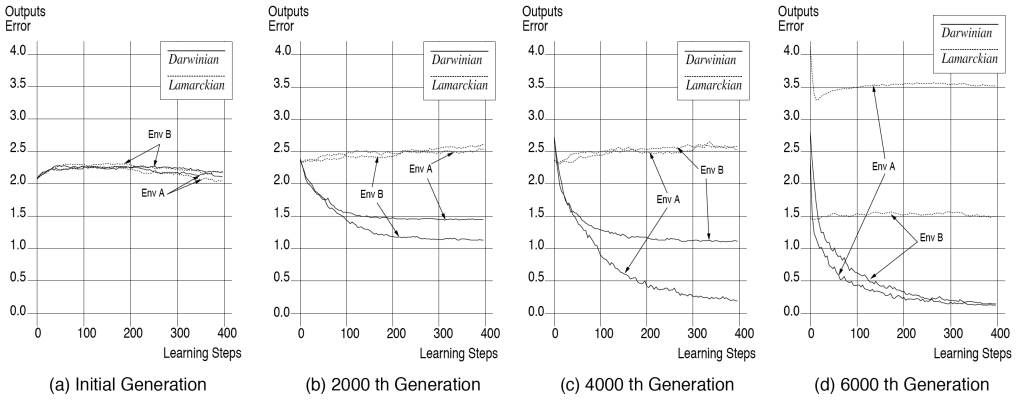

Figure 2.11: Experiment 2 –The change of learning curves across generations

On the other hand, the Darwinian agents succeed in their adaptation across the two environments. In contrast to the first experiment, however, the evolutionary profile is more complicated; the initial rapid increase of fitness is replaced by the slight increase from roughly the 1000^{th} generation. The next phase is more radical; a wide-range oscillation begins around the 3000^{th} generation. This continues for 1700 generations, and then fitness quickly converges on the highest point of the oscillation. After the convergence, it continues its gradual optimization, though it still mildly oscillates. Although the exact reason that the evolutionary process shows this type of profile is unclear, at the end of the experiment, the Darwinian agents attain almost the same fitness value as in the first experiment.

As in the first experiment, Sasaki & Tokoro conduct a small parallel experiment to check how learning is improved along the generations. Figure 2.11 (p. 2.11) reveals an interesting fact; while the Lamarckian agents behave poorly across the generations, the Darwinian agents gradually adapt to the world. More importantly, however, in this experiment their initial error rate never drops; the Baldwin canalizing effect does not take place in the experiment. This is confirmed in further experiments (*see* Sasaki & Tokoro (1998) and Sasaki & Tokoro (1999) in which six types of environment nonsensically change, and under the environments, rules also change). Sasaki & Tokoro interpret the result as the consequence of increasing the "ability to *learn* the task," rather than "ability to *perform* the task". They assume that this is another factor that confines the canalizing effect.

However, this is a somewhat misleading assumption. The "ability to learn" is certainly improved over the generations. This is undeniably a case of genetic evolution. In other words, some part of the task loaded on learning are replaced by genes so that they provide a better configuration of connective values which is ready to cope with both types of environments. However, the very reason that the initial error rate is high is, once again, most likely attributed to the nature of neural network architecture; it may be impossible to configure all connection weights so that the network can perfectly cope with the diagonally different tasks from the beginning; along with the immediate emergence of the Baldwin expediting effect, even if evolution fails to canalize specific knowledge of instances which learning has accomplished, the learnt knowledge is generalized and hypothesized innately. Thus while under a specific environment, an initial error rate would be high, it immediately drops because of the generalization effect. Although Sasaki & Tokoro term this increase of "ability to learn", from a different point of view, this should be considered as a case of the categorizing effect; the learnt knowledge is *indirectly* canalized.

In the other two articles (Sasaki & Tokoro 1998, Sasaki & Tokoro 1999), Sasaki & Tokoro investigate effects of different *heredity rate* of learnt knowledge. If the rate is set to 0, it is a perfect Darwinian inheritance system, while the value 1 designates a perfect Lamarckian system. They conduct experiments with various transmission values, and confirm similar results to those shown above.

2.8.5 Socio-Cultural Learning and Individual Learning –Best

Richard Belew (1990) attempted to consider a new socio-cultural factor going hand in hand with the Baldwin effect. More precisely, he introduced inter-generational information transmission instead of conventional 'individual learning'. The result was quite positive; he showed that the orthodox Baldwin effect could be replaced by a socio-cultural factor. Subsequently, inspired by his works of inter-generation transmission scheme, Michael Best (1999) investigates an intra-generation transmission system.

In both studies, the socio-cultural transmission mechanisms are implemented as 'imitation'; selected agents act as 'models' and other agents copy these models' phenotypes. Interestingly, Best assumes that this type of learning scheme does not include the idea of trial-and-error involved in learning; rather he reckons that imitation is a more reliable, more error free learning scheme than the individual learning scheme. Thus, all 'student' agents can obtain the models' phenotypes with 100% accuracy. In other words, socio-cultural transmission in Best's simulation is cost-free.

In the inter-generational model, model agents are acting as adults, and broadcast their information to their offspring. On the other hand, the intra-generational learning takes place within the same generation. We can reckon these two different learning schemes to have distinctive actual social modes; the inter-generational learning is somewhat close to educational information transmission in juveniles while the intra-generational learning can be considered as a part of life-long cultural information transmission.

As it appeared in the title "*How culture can guide evolution*", Best is primarily interested in seeking to answer the following question: Can cultural influences guide evolution in the absence of individual learning? The method is, again, modeled on Hinton & Nowlan (1987). In the simulation, a type of fitness is introduced to select individuals as 'models'. Then models select learners and they learn models' phenotypes perfectly (*i.e.*, error free learning). The fitness function is described as follows; first it differentiates winners and losers. Winners are those who have all $\boxed{1}$ alleles in their phenotype (without learning). The rest of the population are

all losers. Winners are assigned a sort of fitness value α and losers receive $\alpha/500$. These values correspond to the number of learners the individuals can teach when they act as models. Thus if the value of α is 1, each winner can teach one learner in the population, while losers can teach with only 0.2% chance. In other words, if the population does not include any winners, only two individuals can learn ($1000 \times 0.002 = 2$). Therefore, the value α parameterizes the transmission force. This function produces a snowball effect, as the number of winners increases, more models can teach others.

First, Best conducts a simulation without individual learning. When the value of α is low (*e.g.*, α=5), the social learning cannot support the genetic algorithm to gain a handle on the adaptive goal. However, when the value hits 10, we see a phase transition; [0] alleles sharply drop their frequency in the genepool. For α=20, the shape of transition is close to that of the individual learning. Above that point, the social learning algorithm outperforms the individual learning algorithm. Best also conducts the simulation with the combination of individual learning; the result is similar to the simulation without it.

However, regarding the number of [?] alleles, the social learning algorithm cannot do much about it. In other words, the algorithm succeeds to produce the expediting effect, but fails to offer the canalizing effect. This is for a somewhat obvious reason; the social learning algorithm does not include any errors. Therefore, it does not produce enough pressure to reduce the [?] alleles.

Based on the result, Best also gives an analytic account of the advantage of the transmission system. He mathematically proves that the socio-cultural learning scheme is more efficient than the individual learning scheme. In the simulation, any individual who includes [0] alleles in her genotype has no chance of being a winner. On the other hand, if the individual includes *only* [1] or [?] alleles, she can be considered as a *potential* winner. Based on this, first Best calculates the possible number of potential winners in the initial population whose size is 1000; roughly three potential winners are usually in the population. For α=100, the probability that the three potential winners become real winners in 1000 rounds of social learning is roughly 0.035, while under the same condition, the individual learning algorithm produces the probability of 0.018. Thus the social learning algorithm roughly doubles the chance. Best also converts this result to the learning bias in the individual learning algorithm. In Hinton & Nowlan, [?] alleles express either [0] or [1] equally. To equalize the effect of the individual learning algorithm to that of the social learning algorithm, the individual learning algorithm has to select [1] with 51.4% probability, and 48.6% probability for selection of [0].

Finally, Best investigates a situation where the two learning algorithms plus genetic evolution work based on two different objective functions. He conducts a simulation where the individual learning algorithm and the evolutionary search system are optimizing for the all [1] adaptation goal, while the social learning algorithm is optimized for the all [0] configuration; a diagonally different phenotype. Interestingly, even α=600, genetic evolution with the individual learning eventually wins out over the social learning. At the value of α=900, one winner can teach almost the rest of the population, and genetic evolution loses its foothold. One possible interpretation of this interesting result is that while it is easy for the social learning algorithm to guide evolution, it is rather hard to modify the frequencies of alleles in the genepool.

–Cangelosi & Harnad

Recently, Cangelosi has conducted a series of neural-network-based simulations in which he models a virtual mushroom world. In one of his simulations, Cangelosi & Harnad (2002) argue for the importance of social learning together with individual learning in the formulation of knowledge of categories. They metaphorically describe individual learning and social learning of acquisition of the knowledge as "*Sensorimotor Toil*", and "*Symbolic Theft*" strategies, respectively. These analogical terms capture the differences in the two learning modes. First, in contrast to the Toil strategy, the Theft strategy can circumvent errors which are often yielded on the Toil strategy. In the simulation, mushrooms are either edible or poisonous. Through the Toil strategy, individuals may sometimes take poisonous mushrooms. On the other hand, individuals who take the Theft strategy greatly reduce the risk by relying on the information acquired by individuals' experiences of trials and errors (*i.e.*, through the Toil strategy). This means that the cost of social learning is generally much smaller than that of individual learning. This is along the lines of the result of Best; in the simulation, the number of [?] alleles remains, while the population attains the maximum fitness. Lack of cost of learning in the social learning algorithm accounts for this fact.

Secondly, social learning can take place in a place isolated from the origin of the information. In the simulation of Cangelosi & Harnad, social learning of differentiating mushrooms is done not in front of mushrooms, but in a different place. This is especially true, if the learning is done symbolically.

2.8.6 Mayley

With his series of studies, Mayley makes important observations about the nature of the Baldwin effect in the context of learning and its cost (Mayley 1996*a*, Mayley

1996*b*, Mayley 1997). This section presents three issues relating to the Baldwin effect.

Firstly, Mayley speculates on the costs of learning and their effects on canalization. In his paper (Mayley 1996*a*), Mayley lists the following four types of learning costs. The first is 'time-consumption' learning cost. The individual has to spend some period of its lifetime acquiring certain behavior or a physical trait which can be avoided if the individual is equipped with them innately. During this period, infants typically require special care, such as parental care. Or, if the individual spends too much time acquiring the trait (or, say, maturing), it directly means that the individual loses its reproduction and residual times.

The second type of learning cost comes from incorrect behavior. If an individual performs inappropriately due to inadequate or incorrect acquisition of target behavior, it would decrease its own fitness. In this regard, the learning is costly for the individual.

The third is called 'genetic complexity'. To express the regulatory processes which occur in any structural modification along with learning, generally requires more complex genotypes than simple, innately specified traits. Co-ordination of such complex genotypes would be more susceptible to mutational disturbance. This fragility can be considered as a cost of learning.

Finally, he points out the cost of the learning process itself. An individual will expend its energy while looking for the most appropriate behavior; this would be a vital cost if, for example, nutritional supplies are sparse. Note that while the 'time-consumption' type of cost of learning literally refers to "the cost of time", this type of cost of learning refers to the cost of the activity itself. Thus, although the cost of activity has strong correlation with consumed time of the activity, this consumed time should be separately argued from the first type of cost of learning; it should be considered as a part of the cost of learning itself.

Based on his classification of these learning costs, Mayley argues that genetic specification of behavior or physical traits; innately specified traits are preferable to learning. In other words, under a costly learning circumstance, the canalization process would take place so that individuals reliably reduce the learning costs. By modeling two abstracted learning costs (*i.e.*, explicit, and implicit learning costs) in computer simulations he tests his assumptions. In the first simulation, he investigates the relationship between explicit learning costs and the emergence of the canalizing effect where the learning cost is independent from the learning process itself. Mayley names this as "*posthumous*" learning cost since it is evaluated after the learning process is completed in the computer simulation. Note, however, that in the real world, this does not necessarily denote that the cost occurs after the

end of a learning process. Rather, this shows that the cost is explicitly separable from the learning itself. In this sense, the two types of learning cost –incorrect behavior and genotypic complexity would correspond to this explicit learning cost. The result of the first simulation clearly shows that the high value of learning cost derives from the rapid pace of canalization.

In the next simulation, the implicit learning cost is tested. Note that no factor of learning cost appears in its fitness calculation. This means that learning cost is inseparable from the learning process itself. The time-consumption and energy-consumption learning cost would be categorized as this cost. The learning cost emerges from the difference between an innately predisposed individual's and a learning individual's learning fitness values, if they are cumulative. If an individual has innately predisposed behavior, the fitness value will always be optimal during its whole life, while a learning individual might have a more fluctuating fitness value before it reaches its optimal fitness. Therefore, the learning individual always possesses less fitness value than the innately predisposed individual can possess. Interestingly, Mayley reveals that this implicit learning cost causes a more rapid pace of canalization than the explicit learning cost does.

Finally, Mayley conducts a simulation under a no-learning-cost circumstance. Not surprisingly, in this case, the canalizing effect is completely suppressed. Consider, for example, that a population tries to reach the top of a Fujiyama landscape. Since under the no-learning-cost circumstance, there is no restriction for any inappropriate behavior[8], every agent can try to reach the top forever until he actually reaches it. Obviously, there is no penalty even if an agent reaches a wrong position on a fitness landscape –he can reset the trial completely and start once again. In this situation, canalization has no advantage whatsoever. Since a canalized agent, which incidentally appears in the population by mutation or recombination, does not have selective advantage over the other agents, he cannot disproportionately expand his offspring in later generations. Subsequently, we will not observe the canalizing effect in the populational evolution. Recall that the result is basically the same as Best's (1999) social learning model. In summary, Mayley points out that learning cost is the crucial factor of the canalization process.

Although his first studies show the emergence of the Baldwin effect, prevention of the Baldwin effect is discussed and tested in his other two studies. First, Mayley (1996*b*) reports that the Baldwin effect occurs properly only when the two fitness search mechanisms –evolutionary and learning mechanisms correlate (*i.e.*, G-P correlation). He names this "*neighborhood correlation*". Suppose that the two search

[8]Even if there is no restriction on the number of trials; if there is, it turns out to be an implicit learning cost.

mechanisms have different fitness landscapes. This is as if two search mechanisms stand on completely different landscapes. Suppose also that an agent has a good phenotype that enables the agent to occupy a higher position on the landscape of learning search. Since the agent becomes prolific, it may reproduce a large amount of offspring; it sets its 'base-camp' on the evolutionary search landscape. If the two landscapes are the same, any mutation or recombination that enables its offspring to be closer to the top reduces the burden of learning. The closer the offspring is to the top, the less learning he has to do. However, if the two landscapes do not correlate, hillclimb movements from the base-camp on the evolutionary search landscape do not necessarily help hillclimb movements in the learning search landscape. In this case, even if learning effectively finds the global optimum, evolution cannot follow the path since any phenotypic change cannot be reflected in evolutionary genotypic change. This factor takes quite an important role in the Baldwin effect since his result of high cost learning still fails to derive the Baldwin effect efficiently. This deteriorating effect of "*genotype-phenotype decorrelation*" (**G-P decorrelation**) brings us to an important reconsideration of the Baldwin effect. This will be discussed in Chapter 6, and experimentally examined with the simulations of language evolution in Chapter 9.

Secondly, Mayley (1997) observes a populational prevention of the Baldwin effect called "*the hiding effect*". The hiding effect is another blocking effect of the canalization process. Suppose that there are considerable variances of genotypes in a learning population. If learning enables individuals to acquire the same trait reliably, the genotype's differences become irrelevant to natural selection. In other words, even if individuals have considerable differences in their genotypes, natural selection would fail to tell the differences because their fitness values would be the same by virtue of the learning search. Consequently, the canalization process is suppressed in the population. The term "hiding" is used because of this phenomenon; *hiding* from natural selection. Finally, based on his simulations, Mayley demonstrates using an example that the same phenomena can be observed in an environment in which a fitness landscape is highly rugged. These results are similar to the results under the low cost learning circumstances; in both cases canalization is significantly weakened after individuals can reliably reach the optima. However, there are significant differences between the two. The hiding effect has its root in genetic variations in a population scale while the value of learning cost is one of the environmental factors which are independent of the population itself. Fundamentally, this is the same claim as Deacon's (2003) concept of the "masking effect". A brief description of this concept will be discussed in Chapter 10.

Chapter 3

Language Evolution & The Baldwin Effect

3.1 The Theoretical Approach

3.1.1 Waddington

A year before his passing away, Waddington left a short essay, *The Evolution of Altruism and Language*. In the essay, he made reference to language evolution after describing recent studies of the evolution of altruism and convention. Originally this essay was unpublished, but a year later Waddington (1975) included it in the last section of his edited book. Interestingly, in this essay he already argued that language evolution is based on the Baldwin effect (in his term, "genetic assimilation"). Furthermore, he dealt not with evolution of language itself, but with linguistic ability. These facts are truly remarkable as these basic assumptions he made almost 30 years ago are the assumptions most commonly shared by current researchers and are forming crucial foundations of their studies. As this has been virtually neglected among scholars of language evolution, it is worth sparing a section to introduce this short essay.

The theme of this essay is the evolution of social behavior (especially in human and other higher primates' societies). As is evident from the title, Waddington first argued from altruistic behavior. Evolution of this type of behavior often invokes the necessity of explanation on the population level. While he did not provide his original explanation of altruism, he took up studies of group selection, and kin selection.

Another type of social behavior which is considered as a case of evolution on the population level in the essay is conventions. Males in many species often fight each other for territories, breeding females, or food resources. Although such competitions often involve actual fighting, those competitors do not usually reach the point where they suffer critical injuries. This is due to the fact that the losers display behavior which is recognized by the winners as 'throwing in the towel'. Waddington considered it as a sort of socially recognized convention. Maynard Smith's (1982) attractive theory of "*Evolutionary Stable Strategies*" (**ESS**) was just introduced

at that period, and Waddington introduced it to claim that conventional behavior is partially explained not on the population level, but on the individual level. While he mainly accepted the theory of ESS with a positive attitude, Waddington also argued against it. He concerned that while it provides an attractive explanation for strategies which are evolutionarily stable, no explanation for why the population reaches specific strategies is given in the theory. Regarding this 'defect' strategy, Waddington posed the two following questions. The first question is about ubiquitousness. For a convention to be advantageous, a non-trivial fraction of the population has to accept it. However, it is obvious that such conventional behavior has to begin at the individual level; there must be a period in which a small fraction of the population had exercised such behavior. The ESS does not tackle this question. The second question is how, among possible behavior, a specific type of behavior has been selected as a convention. This is also not addressed in ESS. Interestingly these questions strike exactly the main points which Brian W. Arthur (1994) and other 'complexity-oriented' economists have addressed (this point will be briefly discussed in Chapter 6).

After discussion of these questions, Waddington started arguing about language evolution. Initially, he described it as the most complex example of convention.

> The selection of an item of behaviour to act as a conventional sign, part of a system of communication with another individual, can perhaps be regarded as a first evolutionary step towards one of the most complex and certainly one of the most important of all social characters: the ability to use language.
> (Waddington 1975, p. 304)

This statement serves as a good reflection of his basic stance towards language evolution. First, he clearly related language evolution to the emergence of socially agreed, conventional behavior. Secondly, this emergence problem is considered not directly through the emergence of this conventional behavior itself, but through evolution of 'the ability to use language'. Based on this, Waddington stressed the gap found between human beings and other animals. He argued that genetic evolution of this ability must be gradual as opposed to Chomsky's claim of a macro-mutation (*e.g.*, Chomsky 1972, Chomsky 1982*a*, Chomsky 1982*b*, Chomsky 1988). Furthermore, as an epigeneticist, Waddington cautioned mutations often cause minor or no phenotypic effects except in cooperation with certain particular environmental factors. Then the gap –no intermediate state in linguistic ability is a mystery. Nevertheless, he claimed that the epigenetic theory might help to solve this problem; Waddington proposed a theory based on a principle, called "the principle of archetypes". This

is fundamentally the same as the "punctuated equilibrium" theory of Stephen J. Gould and Niles Eldredge (1977). Basically, this theory states that accumulation of small modifications suddenly produces a massive difference. In the case of language evolution, this produces a significant degree of conformity, and consequently other intermediate stages have been rapidly supplanted:

> Language which developed to the human state may well be such an archetypal novelty. It would be such an incomparably more effective means of communication than any system which had a few but not all its major properties, that any such evolutionary intermediates would have been very rapidly supplanted; in such circumstances one could hardly expect to find anything of them surviving to the present day.
> (Waddington 1975, p. 306)

Waddington suggested that this archetype theory does not require any 'macro-mutation'. The effectiveness which he discussed here, is not caused by modifications on hereditary factors but arises from epigenetic development, namely learning. Thus this is a candidate for his theory of the Baldwin effect. He assumed that however primitive, any rudimentary language is selectively advantageous. If language use is important, learnability itself also becomes important. This leads to the following statement:

> If there were selection for the ability to use language, then there would be selection for the capacity to acquire the use of language, in an interaction with a language-using environment; and the result of selection for epigenetic responses can be, as we have seen, a gradual accumulation of so many genes with effects tending in this direction that the character gradually becomes genetically assimilated.
> (Waddington 1975, p. 306)

Later, Waddington more explicitly asserted that a kind of LAD is the target of genetic assimilation (*i.e.*, in the sense of canalization):

> [B]ut rather his mind contains certain rather definite capacities for handling symbolic communications systems of a particular kind in particular ways. It is this particular mental apparatus which I suggest might have been built up by a process of genetic assimilation.
> (Waddington 1975, p. 307)

3.1.2 Pinker & Bloom

It would not be too much to say that Steve Pinker and Paul Bloom (1990) liberated the discussion of the evolution of language. Until 1990, the study of evolution of

language had remained as an unattractive subject. This is perhaps largely because Chomsky has a dismissive attitude towards the evolution of language. Chomsky has emphasized that evolutionary theory is not so informative about the question of the evolution of language. He believes that Darwinian theory has little to say about the origin of language or the course of its evolution. This strong position against studying the evolution of language is twofold.

The first point comes from the denial of functionalism. Chomsky claims that the explanation for fundamental properties of language could be found not in its functional aspect but in its forms. This directly follows from his skepticism towards an adaptive account of language evolution[1]. If language form is independent from fitness, how can natural selection shape the current forms of language? Besides, language forms are quite distinctive from other complex biological or cognitive systems on a number of points. For instance, he discusses the redundancy in biology and language forms:

> [I]t has often proven to be a useful guiding intuition in research that if some property of language is "overdetermined" by proposed principles, then probably the principles are wrong, and some way should be found to reconstruct them so as to avoid this redundancy... Typically, biological systems are not like this at all. They are highly redundant, for reasons that have a plausible functional account. Redundancy offers protection against damage, and might facilitate overcoming problems that are computational in nature. Why language should be so different from other biological systems is a problem, possibly even a mystery.
> (Chomsky 1991, pp. 49-50)

The second point is his skepticism about natural selection regarding the computational properties of the brain. He argues that the evolution of language has happened through the evolution of highly concentrated brain structure. However, this process is not explicable by the theory of natural selection, he claims.

> Perhaps these [properties of language] are simply emergent physical properties of a brain that reaches a certain level of complexity under the specific conditions of human evolution.
> (Chomsky 1991, p. 50: emphasis by author)

[1]Note, however, that Chomsky never denies that the communicative aspect of language contributes to improve the adaptive advantage of human beings. What he is skeptical about is the idea that communicative aspects can thrust language to its current form.

> We know very little about what happens when 10^{10} neurons are crammed into something the size of a basketball, with further conditions imposed by the specific manner in which this system developed over time.
> (Chomsky 1975, p. 59)

This strong attitude has greatly dissuaded a number of linguists from engaging in the study of language evolution.

By emphasizing that the only possible algorithm for language evolution is natural selection, Pinker & Bloom cast a strong doubt on Chomsky's claims. There are two major issues in biology loosely supporting Chomsky's claims –the 'spandrel' theory and the 'exaptation' theory. Pinker & Bloom argue that language is explained by neither theory, and has evolved gradually by natural selection. In this sense, their arguments are based on a very biological and conventional wisdom. Indeed, they suggest that the Baldwin effect may be involved during the course of language evolution. More specifically, Pinker & Bloom ask the same question as Waddington asked. That is, given a mutated individual whose grammatical ability is higher than other extant members, how could such a person be more adaptive, if he only possesses a better communicative envelope? Pinker & Bloom point out that even in modern communities, we can find some discrepancies between the abilities of utterance and comprehension; human beings are often able to comprehend ungrammatical utterances. Also, some are better speakers whose expressions have never been expressed by others. Yet, others can appreciate such 'new' expressions. Consider, for example, the case of Shakespeare; although almost nobody could innovate such sophisticated expressions, they have been acclaimed because others can comprehend them. However, Pinker & Bloom assume, to comprehend such novel expressions, unprecedented cognitive efforts are required. Then this becomes a selective pressure. Eventually, this pressure triggers the Baldwin effect. Pinker & Bloom state:

> When some individuals are making important distinctions that can be decoded with cognitive effort, it could set up a pressure for the evolution of neural mechanisms that would make this decoding process become increasingly automatic, unconscious, and undistracted by irrelevant aspects of world knowledge. These are some of the hallmarks of an innate grammatical "module" (Fodor 1983). The process whereby environmentally induced responses set up selection pressures for such responses to become innate, triggering conventional Darwinian evolution that superficially mimics a Lamarckian sequence, is sometimes known

> as "the Baldwin Effect".
> (Pinker & Bloom 1990, p. 722)

Nearly fifteen years after their paper, however, current research agendas are starting to look in a slightly different direction. While Pinker & Bloom emphatically discuss that the only possible explanation of language evolution is biological one, they seem to consider phylogenetic evolution almost exclusively. However, as the Baldwin effect considers, learning provides a different mode of evolution –yet, it is *purely* biological. It is somewhat unfortunate that even though both linguists are strongly motivated by generative linguistics which sets language acquisition (thus, learning) as the central issue of the theory, they fail to recognize a possible evolutionary role of learning itself in their theory of language evolution.

3.1.3 Deacon

As we will see in Section 4.2.3, in *The Symbolic Species*, Deacon greatly contributes to opening a new avenue for the Baldwin effect, in this section, other points derived from the book are briefly discussed.

Firstly, as a renowned researcher in neuroscience and evolutionary anthropology, Terrence W. Deacon puts more stress on the biological plausibility than any other researchers do. It is generally true that researchers in other fields loosely define genetics in their models, especially the relationship between genotype and phenotype. However, Deacon reckons that all behavioral traits should get neurologically plausible supports. This is somewhat similar to Waddington's attitude towards the Baldwin effect. Therefore, even if a theory is linguistically appealing, if it lacks a neurological foundation, he will not buy the argument. This point becomes the most salient in the study of UG regarding the Baldwin effect.

In one chapter of *The Symbolic Species* (Deacon 1997), he speculates on the evolution of language with regard to the Baldwin effect. While several related topics are addressed in the chapter, some important claims that he makes are cited here. First, he points out that the general and fundamental features of language, namely UG, have to be persistent for hundreds of generations.

> The relative slowness of evolutionary genetic change compared to language change guarantees that only the most invariant and general features of language will persist long enough to contribute any significant consistent effect on long-term brain evolution.
> (Deacon 1997, p. 329)

Deacon also claims, however, that this argument cannot coexist with the idea of UG in generative grammar. He asserts that, to be canalized, language has to be

processed in the same neurological regions of the brain regardless of language or person. In addition, this has to be reliably done throughout a considerably large number of generations. The problematic point of the account of UG in this regard are the various surface implementations of the grammars in different languages:

> The very abstraction from the surface implementation of morphology and syntax that provides the grammars with their generative power also shields [aspects of the deep grammatical logic of language] from the reach of natural selection.
> (Deacon 1997, p. 339)

He concludes:

> Therefore, they are the *least* likely features of language to have evolved specific neural supports. Those aspects of language that many linguists would rank most likely to be part of a Universal Grammar are precisely those that are ineligible to participate in Baldwinian evolution!
> (Deacon 1997, p. 333: original emphasis)

Since natural selection cannot see any types of linguistic traits which bridge all natural languages, innate properties of language, general linguistic constraints, and syntactic categories are not subject to the Baldwin effect, Deacon argues. Then any kind of canalization cannot take place. This leads him to the final conclusion:

> No innate rules, no innate general principles, no innate symbolic categories can be built in by evolution.
> (Deacon 1997, p. 339)

Clearly, this is an extremely strong claim. The most important point in his claims is that theories of the current linguistic studies do not have any neurobiological foundation whatsoever. This directly opposes the idea of the LAD; he criticizes the theory of the LAD as "*monolithic innatism*" (Deacon 1997, p. 350). Nonetheless, Deacon claims that languages must have evolved hand in hand with the evolution of the brain by virtue of the Baldwin effect. At a glance, he seems to be in deep trouble; Deacon is now asked to explain the evolution of language without citing universality of natural languages. However, he circumvents this problem by putting forward a co-evolution theory between linguistically independent cognitive abilities and the brain structure. In this sense, his argument surely supports a no-innate-but-pure-learning theory of language acquisition. He considers that language acquisition is supported not by monolithic innatism –language acquisition *par excellence*, but various cognitive predispositions, which are almost irrelevant to language acquisition on their own. The Baldwin effect can support the development of simple predispositions, such as cognitive abilities of attention, imitation,

or automatic reflection. If these predispositions, *as a whole,* enhance fitness during one's lifetime, it will demand the brain to enhance its performance. Since such predispositions are simple and sufficiently environmentally universal, they might be subjects of canalization. Consequently, it enhances the further predispositions and incorporates much broader regions' adaptation. It forms a cycle of co-evolution of languages and the brain; a robust, failure-free mechanism of language acquisition. The trick in his argument is that the Baldwin effect does not work directly on the language faculty. Rather, it affects other behavioral predispositions.

Together with Waddington's attitude towards the Baldwin effect, Deacon's requirement of a neurobiologically plausible account in the evolution of language gives a serious take-home problem for all researchers in this field. Currently, a number of researchers make a rather straightforward relationship between a linguistic learning mechanism and its representation of genotypes. However, such theorists are highly skeptical of this assumption; linguistic phenotypes might be too complicated for genes to express. Deacon's reduction to a more genetically simple behavioral account has to be carefully considered.

3.1.4 Dor & Jablonka

Recently, a linguist Daniel Dor and Eva Jablonka published two articles regarding language evolution (Dor & Jablonka 2000, Dor & Jablonka 2001). The two articles are almost identical; in the papers, Dor & Jablonka deploy their model of language evolution based on Dor's linguistic theory and Avital & Jablonka's (2000) extension of the Baldwin effect (*see* Section 2.7). Treating the articles as one, in this section, we look into how the synergy of the two theories works and what type of evolutionary model is presented.

In the papers, first Dor & Jablonka argue against basic tenets of two major camps in modern linguistics. They claim that under the light of language evolution, both of the two main camps of linguistics, namely *formalism* and *functionalism*, reveal their fundamentally flawed assumptions. Regarding the degree of domain specificity of linguistic abilities, both camps' claims are equally extreme, but in opposite ways. While formalists (*i.e.*, generativists) strongly stress the idiosyncrasy of linguistic abilities, and claim innate modularity, functionalists have denied such linguistic properties and have drawn diametrically different conclusions. They have claimed that linguistic knowledge is reducible to less domain-specific cognitive principles. No specific module for language is required.

Regarding formalism, Dor & Jablonka point out the following four problems. The first one is inconsistency with empirical data. By drawing an example from famous linguistic constraints on grammatical extractions in English, called *island*

constraints (Ross 1967), Dor & Jablonka argue that there are a lot of attested cases where noun-phrase embeddedness violates the island constraints. Second, from the evolutionary point of view, it is well known that formalism's exceedingly dysfunctional theory of language does not allow any Darwinian theory to be attainable. The third problem is about the notion of innateness. In formalism, innate linguistic knowledge is highly specific; there are abundant formalists' theories which both explicitly and implicitly assume that specific grammatical rules and constraints are genetically encoded. This strong genetic determinism does not reconcile with what neuroscience or some other related fields tell us. The final problem of formalism is their static model of language. Formalism places one of its principles of methodology on synchronicity. Subsequently, their view of linguistic knowledge is static and universalistic. Outcomes are, Dor & Jablonka claim, highly problematic when we consider evolution of language; they claim that linguistic knowledge is more dynamic and variable.

In the case of functionalism, first of all, Dor & Jablonka argue against its too general attitude to linguistic knowledge; its excessively reductionistic attitude fails to provide tenable explanations for specific grammatical facts. Moreover, they claim that reducing linguistic capacities to general cognition does not match the well-known facts in linguistics –language acquisition, language breakdown, and the formation of *de novo* languages (*e.g.*, Nicaraguan sign language). All these instances seem to suggest that linguistic abilities are to some extent unique, and manifestation of these phenomena may not be attributed to general cognition. Moreover, even if we accept that linguistic knowledge relies on more general cognition, as Marr (1982) claimed, any functional account of cognitive abilities should be much more specific. In this regard, functionalist claims are too general.

Given these criticisms, Dor & Jablonka put their axis somewhere between the two camps; on one hand, similar to functionalism, they claim that linguistic properties are sensitive to the demand of meaning. Thus there is some space for functional explanation in the matter of linguistic properties. However, in contrast to the functionalist school, their claim is much more moderate; while the origin of linguistic knowledge came from the demands of semantics, over the generations, it gradually comes to be genetically encoded. Regarding properties of linguistic knowledge, they deliver their own account of linguistic properties with an example of island constraints (Ross 1967). The core argument in the explanation of the constraints is as follows: While formalists have asserted that island constraints are one of the core linguistic properties which meaning cannot do anything with, they claim that the phenomena are explainable on a semantic basis. They separate "*Event Structure*" and "*Epistemic Licensing*", and, the constraints come from this distinction. Thus

the explicit principle of island constraints is redundant. However, they also admit that the theory is still in the very early stage of its development, so it does not have much explanatory power to cover many linguistic properties.

In summary, according to Dor & Jablonka, the categories, event structure and epistemic licensing, belong to a uniquely linguistic level of meaning representation. The schematic representation of this view is given;

[conceptual representations]

$\updownarrow$

$\|$ [linguistic meaning] $\leftrightarrow$ [linguistic form] $\|$

$\updownarrow$

[phonetic representations]

This representation schema captures their view of language as a transparent mapping system between the levels of "*Linguistic Meaning*" and "*Linguistic Form*". They conceive that the conceptual representations are not only for human beings, but other primates have them, though they may be more rudimentary. And the conceptual representations have different communicative channels –emotional meanings are communicated through facial expression, body language, music etc. Linguistic Meaning is a subset of conceptual representations, and it sets language's expressive envelope. Thus language is described as a communication tool that is structurally designed for the communication of a constrained set of meanings. Dor & Jablonka define evolution of language as evolution of this mapping system[2]. Their claim is that this characterization of language re-frames the question of language evolution. It is neither the evolution of a dysfunctional, formal system, nor the evolution of a general-purpose communication system supported by a number of general cognitive apparatuses. Rather, it is the evolution of a 'specific' communication system, a small subset of the overall communication system.

In the next sections of the papers, based on their theory of language, they deploy their model of language evolution. To bridge the two models, they first provide the following list as a summary of their assumptions.

1. A linguistic theory should be a semantically based and empirically oriented theory of transparent meaning-form relations.
2. However, it should also not be reduced to general cognition.
3. Linguistic ability as a whole is not a general-purpose communication system, nor a formal system. Rather language is a functional,

[2]By saying Linguistic Form, Dor & Jablonka do not provide evidence of exactly what type of linguistic form they describe.

unique, and transparent *mapping system* between the representational level of *linguistic* meaning and the representational level of *linguistic form*.

4. According to this theory, evolution of language is gradual expansion and sophistication of the linguistic mapping system
5. This produces the following three distinctive questions. How did language evolve? How did speakers (and linguistic ability) evolve? How did these two processes interact?
6. Cultural evolution played a major role in the evolution of language.
7. Behavioral plasticity played a crucial role in this evolution.
8. However, differences in the ability to learn (*i.e.*, plasticity) became selectively important from time to time (*i.e.*, this type of selection occurred periodically). This produced linguistically biased cognition.

Based on the theory developed by Avital & Jablonka (2000), Dor & Jablonka develop their theory of language evolution; the evolution of the linguistic mapping-system described above. Dor & Jablonka admit that the process of evolution is multi-faceted, and different questions can be investigated. This includes, for example, inquiry about the type of selection pressures, possible stages of the evolution, or a branching pattern of languages. In the articles, they confine themselves to focus on the dynamic patterns of the evolutionary process.

In the investigation of the dynamic patterns of the evolutionary process, they provide two different modes of evolution, namely cultural and genetic evolution. To bridge these modes of evolution, they introduce the Baldwin effect. In contrast to other researchers, their theory of the Baldwin effect introduced here is strongly reinforced by the study of Avital & Jablonka whose formulation of the theory is more explicit and theoretically well-formed. Dor & Jablonka apply the work to their own study of language evolution.

Similar to their theory of language, Dor & Jablonka take a somewhat intermediate stance in their evolutionary theory; language evolution is a bilateral process of cultural and genetic evolution. To simplify their evolutionary model, they assume language evolution as comprising an arbitrarily long number of stages. They concentrate on two early, consecutive stages, N and N+1. In stage N, hominids in a community are equipped with the necessary precursors for linguistic communication. Their conceptual envelopes are much larger than their expressive envelopes (whatever the channels of expression are). Assume also that the individuals use and acquire their quasilinguistic system. Assume further that the community has

some genetic constitution which supports this system. Crucial for the theory, they suppose that the genepool has enough variability regarding this constitution. The difference between stage N and stage N+1 is that N+1 has a more sophisticated and developed expressive envelope. Moreover, genetic constitution has also changed in this stage so that individuals can comfortably acquire the developed system.

In stage N, some individuals incidentally make linguistic "innovations". Dor & Jablonka assume that the driving force of these innovations is a growing pressure for better communication. They stress the point that to make innovations, no particular genetic foundation specifically designed for the innovations is required. Also, it may be the case that only a small fraction of a population can enjoy their innovative capability. However, a larger group of individuals could learn and understand those innovations. For example, the yam-washing behavior by Japanese macaques on a small island was incidentally found by a young female macaque. Other young macaques, however, also learnt the behavior. This is a similar assumption which both Waddington and Pinker & Bloom made.

Although such innovations may prevail among small groups in the population, propagating across the population is a difficult task. A number of innovations initially shared by a small number of individuals fail to deploy themselves into the whole population. Even very adaptive behavior may disappear. As most linguistic innovations relate deeply to communication, such innovations are only fully appreciated when a certain number of individuals exercise them[3]. Dor & Jablonka assume that establishments of innovations are more likely successful after first learners transmit them to the next generation. This is because youngsters generally have great capability to acquire new things; in cultural evolution, children often play an important role in establishing traditions (in the case of Japanese macaques, only young macaques had exercised the yam-washing behavior for the first six years or so.)

Once such innovations are established, then they may become conventionalized and streamlined by the process of iterated learning. These conventionalized and streamlined traditions become, by themselves, constraints on new innovations. This is conceivable as a case of a canalizing effect. Moreover, these establishments of innovations pose different types of demands on the community itself. Individuals in the community have to acquire the traditions and also comply with them. Also, the traditions might change other social traditions, cognitions, or social relations. Thus established traditions change the environment surrounding the community

[3]This is called frequency-dependent selection. We will come back to this point in Chapter 6.

and produce new selective pressures. This is called "*Niche Construction*" (*e.g.*, Odling-Smee *et al.* 2003).

Up to this point, we have not had to invoke a genetic explanation, while such innovations are constrained within the genetically-based capacity of human beings. However, as more and more innovations become social linguistic traditions, more and more new niches are formed on existing environments (*i.e.*, previous niches). In line with this, cognitive demands become more and more severe. During such a cumulative process, some individuals drop out due to the increase of cognitive demands, while others survive. This winnowing process itself reveals hidden genetic variation in the community; gradually genetic differences in individual learning capacity appear. In their terms, the frequencies of those gene combinations which contributed to easier language acquisition and use increase in the population; the canalizing process takes place. What this process targets is the cognitive capacities most useful for the specific linguistic performance. After the process of canalization, another stage begins; once again emergence of innovations eventually induces canalization.

Based on some specificity of linguistic ability argued above, Dor & Jablonka assume that genetic evolution specifically in linguistic ability most likely takes place. Disregarding genetic evolution of general cognitive abilities which produced language as a byproduct, they cite studies of language acquisition as support for their assumption of an innately-given linguistic foundation. The question here is then how this foundation has been evolutionarily formed, if it is not attributed to strong, less domain-specific constraints on brain development in children, as Deacon (1997) has claimed. They do not provide an answer in the papers.

Dor & Jablonka also present some required conditions for the process to work. First, plenty of genetic variation which is phenotypically visible is required. Secondly, different sets of genes should get involved. This is because due to niche construction, a novel selective pressure may work on different phenotypic properties. Selection existing over several generations is also essential.

In summary, together with the work of Avital & Jablonka, Dor & Jablonka succeed in providing a significant insight for language evolution. As we will see in Chapter 6, the following three points are incorporated in our theory:

1. The Baldwin effect can be related to niche construction in the case of language evolution.
2. Exaptation process can be conceivable in the context of the Baldwin effect if niche construction takes place.

3. The Baldwin effect can take place in a cyclic manner so that some part of the communication ability can be evolutionary enhanced (together with the assimilate-stretch principle).

3.1.5 Newmeyer

Newmeyer has been one of the most prominent formalists in linguistics. However, in contrast to other hard-core formalists, he has attempted to deepen his understandings of other schools in linguistics. Extending his interests towards the study of language evolution in general, Newmeyer also commits himself to speculate about the evolution of language under the formalist scheme. In this section, we introduce his study of protolanguage (Newmeyer 2000).

With a mixture of speculations and theoretically backed-up assumptions, in the first half of his article Newmeyer discusses a possible 'Proto-World' word order. First, armed with current statistical data of the world's language typology, the tendency of language change, and its theoretical explanations, Newmeyer argues that *SOV* word order has been much more typologically predominant. His first assumption is the following: "*SOV order predominates among the world's languages today*" (Newmeyer 2000, p. 372).

It is also known, however, that many previous *OV* languages have changed to be *VO* languages. The frequency of this change is far greater than the reverse, although this does not mean that the reverse is impossible; language changes from VO to OV have been also both attested and reconstructed. Given this fact, Newmeyer propose the second assumption: "*The historical change OV > VO is both more common than the change VO > OV and more 'natural'*" (Newmeyer 2000, p. 373).

Together with the first assumption, the second assumption depicts a somewhat confusing picture. On the one hand, *SOV* is statistically predominant in the world's languages. On the other hand, however, the language order seems to be driven away from the world's languages. From this the following interim conclusion is derived. Thus, "*SOV order was once much more typologically predominant than it is now*" (Newmeyer 2000, p. 375).

While this interim conclusion mainly comes from empirical studies, Newmeyer also attempts to coordinate it with a more theoretical explanation. Adopting Bickerton's (1990) theory of 'protolanguage', Newmeyer discusses a possible syntactic feature of the earliest human language (*i.e.*, the immediate descendant of protolanguage). According to Bickerton, the evolutionary antecedent of human language in the current state would associate nonlinguistic conceptualization of events with rudimentary linguistic representations. This nonlinguistic conceptual representation system is called "*conceptual/thematic representation*" (*see* Jackendoff 1983,

Jackendoff 2002). As a crude communication system, protolanguage may have an interface which maps this thematic representation onto a symbolic system. Newmeyer assumes that this interface assigns overt markers, such as inflectional morphology, to such thematic arguments. He states: "*Protolanguage had thematic structure*" (Newmeyer 2000, p. 375).

However, following Bickerton, Newmeyer also conceives that in protolanguage these arguments would occur in no fixed order. Along with this, it might lack argument structure which states syntactic positions of these arguments. Finally, Newmeyer makes one more assumption: "*Protolanguage lacked quantificational structure*" (Newmeyer 2000, p. 375).

While he admits this is purely based on plausibility, it seems to be a natural assumption about protolanguage: If it did not have argument structure with fixed word orders, any reliable quantification with multiple arguments seems to be unlikely.

After these assumptions regarding protolanguage, Newmeyer makes an attempt to associate it with the interim conclusion; *a possible structure of the earliest human language*. Compared with *SVO* languages, *SOV* languages generally have a smaller number of movement rules. For example, the majority of verb-final languages do not have '*Wh*-movement', while less than a half of *SVO* languages lack it. Moreover, rigid verb-final languages tend to have a small number of motivations for moving elements to argument positions. It is also generally observed that *SOV* languages more directly assign the thematic role of syntactic positions, compared with *SVO* languages. This relates to the well-known typological fact of *SOV* languages:

> **Universal 41**
> If in a language the verb follows both the nominal subject and nominal object as the dominant order, the language almost always has a case system.
> (Greenberg 1963, p. 113)

Indeed, a study shows that among 237 languages, 64% of *SVO* languages have explicit case, as opposed 30% of *SOV* languages (Dryer 1989). Newmeyer briefly gives processing efficiency regarding identifying arguments as a possible reason for this tendency; due to extra cues provided by case, even before a head appears (in an *SOV* language, a head usually appears at the end of the argument structure), thematic roles are uniquely identifiable.

However, this strong correlation of thematic structure and the base structure of *SOV* languages makes these languages have a more indirect means of representing the scope of quantification or some other logical operators. Newmeyer presents

the following examples of *wh*-phrases. In English, α is recognized as an indirect question, while β is a direct question, while both sentences are derived from I. This is due to the position of the logical operator "*who*":

α He was wondering *who* you saw.
β *Who* was he wondering that you saw?

I He was wondering *you saw who*.
II He *you saw who* were wondering.

On the contrary, in an *SOV* language, because there is no movement of *who*, an *SOV* underlying structure (*i.e.*, II) and its corresponding surface structures (*i.e.*, α & β) are identical. This sentence form is, thus, ambiguous. Newmeyer states that to resolve scope ambiguities, *SOV* languages use far more indirect means for signaling scope than do *SVO* languages (such as the placement of special question particles *-ka* at the end of a sentence in Japanese, as opposed to *wh*-movement in English).

Given these facts, he proposes:

> a. SVO languages are 'good at' representing quantification directly, but 'bad at' representing thematic structure directly.
> b. SOV languages are 'good at' representing thematic structure directly, but 'bad at' representing quantification directly.
> (Newmeyer 2000, p. 375)

Adopting Bickerton's hypothesis of the transition from protolanguage to true human language once again, Newmeyer proposes that the transition would be the creation of argument structure which is transformed from underlying thematic structure; emergence of syntactic operations. He speculates that the transition prefers the most processing-efficient way; projecting the basic structure of thematic structure onto argument structure. Again, he seeks the advantage of *OV* order against *VO* in the ambiguity of the types of arguments assigned by heads. If a head is followed by its complement, it is often the case that the type of argument the complement takes is left until the complement appears. For example, the English verb "*break*" takes at least five types of arguments.

On the other hand, if such thematic roles are explicitly expressed by case or some morphological, phonological cues, placing complements after verbs is an extra operation. This assumption is naturally drawn from the more direct relationship of thematic structures and surface forms of *SOV* languages. Thus Newmeyer finally reaches the following conclusion in the first half of the article: "*The earliest human language had rigid SOV order*" (Newmeyer 2000, p. 379).

He claims that this conclusion sheds significant light on language evolution. Importantly, in rigid *SOV* languages, the major UG constraints proposed in the Government-Binding framework (Chomsky 1981) are rarely manifest, while in *SVO* languages these are quite common. For example, although the Subjacency constraint, the C-Command constraint, and the Empty-Category Principle are well exemplified among the *SVO* languages, it is quite often the case that the rigid *SOV* languages lack evidence for these UG constraints. In other words, fundamental syntactic rules that regulate structures of *SVO* languages are 'invisible' in *SOV* languages.

Then Newmeyer considers that this fact brings about a suspicion of Baldwinian explanations of language evolution; as found in the studies of Mayley, the Baldwin effect generally takes place when learning is costly. Since the cost of language acquisition failure would be quite expensive regarding one's adaptation, it is a plausible assumption that the Baldwin effect would have taken place to reduce the danger by innately prespecifying those UG constraints; initially children had to learn those constraints from scratch. After many generations, such learned constraints should have been canalized.

The important assumption is this; given the uneven distribution of the major UG constraints in the current world languages according to typological differences of word-order, the emergence of such constraints would be expected to occur along the side of the development of *SVO* languages. However, if such constraints have indeed been canalized by virtue of the Baldwin effect, it would also be expected that the Baldwin effect had taken place in the domain of *SVO* languages but not in the domain of *SOV* languages. There is no reason to expect that the populations of *SOV* languages have undergone the same selection. If such constraints are not canalized in populations whose languages are *SOV*, how come children in these populations can equally learn any *SVO* languages that require innate linguistic knowledge of the UG constraints? Universal learnability of world languages is one of the most fundamental premises of current linguistics. Therefore, the origin of the UG constraints as a canalization scenario holds a serious contradiction. For this reason, Newmeyer draws the following final conclusion.

> UG constraints must have appeared contemporaneously with the appearance of true human language, or they cannot be innate at all.
> (Newmeyer 2000, p. 384)

However, it is somewhat apparent that this logic holds only when we accept the strong assumption that the UG constraints indeed exist, and they are the direct target of canalization. As the claim is purely theory-laden, it would be highly

possible that Newmeyer's point is not be problematic at all. As such, the legitimacy of the claim should be discussed independently.

3.2 The Computational Approach

In this section, we look at recent studies that utilize population dynamic systems in the study of language evolution.

3.2.1 Turkel

Based on Hinton & Nowlan's simulation described in Section 2.8.2, Turkel conducts an experiment that holds a population dynamic communication system. While Turkel mostly adopts Hinton & Nowlan's genetic encoding method (fixed, and plastic genes), he provides an external motivation for it by incorporating Chomsky's (1981) "*Principle and Parameters approach*" (**P&P**). Turkel considers the fixed alleles (*i.e.*, $\boxed{0}$ and $\boxed{1}$) as "*principles*", and the plastic alleles (*i.e.*, $\boxed{?}$) as "*parameters*". Then, the genotype is considered to be the LAD. Since a replicated experiment will be conducted in Chapter 9, only a brief description of his study is provided here.

If one wants to utilize a GA in any type of simulation, one has to design a representation of inheritance. Recall that Hinton & Nowlan adopt a basic binary representation of genes with a small trick; they introduce a third allele, namely a plastic allele. Turkel combines this representation method with a theory of language acquisition. Therefore, he provides an appealing model with which we can test evolution of the LAD by virtue of the Baldwin effect.

The second point is the interactive aspect of his simulation. In Hinton & Nowlan, an agent is insulated from other agents; there is no interaction between any of two or more agents. Recall that in Hinton & Nowlan, plastic alleles express the two types of phenotypic values ($\boxed{0}$ or $\boxed{1}$) while fixed alleles can only express the same type of phenotypic values (*e.g.*, if an allele is $\boxed{1}$, it can express only the phenotypic value $\boxed{1}$). Whether or not an agent succeeds in reproducing depends solely on the fixed objective function. All agents have 1000 chances to modify their phenotypes and these trials are regarded as learning. Turkel modifies this model so that protocol conformity becomes an important factor; instead of comparing agents' phenotypes with the objective function, fitness is measured by the similarity of two agents' phenotypes (it is called 'the Hamming distance'[4]).

[4]The Hamming distance is measured by locus-by-locus based differences of two genotypes. Thus the value of the Hamming distance in two N-long genotypes ranges from 0 (identical) to N (no commonality). In this thesis, this concept is also extended to the phenotypic level as long as their representations are fundamentally retained.

More specifically, in one generation, all agents are selected serially from 1 to 200 (the size of the population is 200 in this simulation). A partner is randomly selected from the same population. Then the two agents try to establish communication by aligning their phenotypes (*i.e.,* grammars); plastic alleles (*i.e.,* parameters) are modified to either [0] or [1] randomly. If the two agents succeed in having an identical phenotype, it is regarded as establishment of communication. Only a perfect match of the two phenotypes can assign a high fitness value; 'similar phenotypes' have no meaning in this fitness function –a needle-in-a-haystack search. As with Hinton & Nowlan, a cost of learning is introduced; the fewer trials, the more fitness value the selected agent can obtain (its randomly selected partner is not assigned a fitness value). If the two agents have a discrepancy on loci which are occupied by fixed alleles (*i.e.,* principles), they cannot establish a communication since there is no chance of aligning values expressed from those fixed alleles. Therefore, the whole population quickly converges on a small number of genotypes so that agents can reliably increase their phenotypic conformity. At the same time, because of the learning cost, plastic alleles are canalized –the canalization process saliently appears in Turkel's simulation.

Since there is no external factor to define the best phenotypes in the population, the optimal configuration of a phenotype is determined by a given dynamic system itself. In other words, the objective function of the model is not fixed. This is also one of the significant differences in Turkel's simulation; unlike Hinton & Nowlan, there is no fixed, arbitrarily decided configuration of phenotype to obtain the highest fitness value. Turkel points out that this dynamic and indeterministic aspect provides a counterargument against the conventional skepticism towards the study of language evolution; since forms of natural language exhibit a number of nonfunctional aspects, evolution of language may not be explained by virtue of natural selection which can 'see' only functional aspects. For instance, Turkel quotes Piatelli-Palmarini's statement: "*Adaptationism cannot even begin to explain why the natural languages that we can acquire and use possess these central features and not very different ones.*" (Piatelli-Palmarini 1989, p. 24). Turkel argues that if such dysfunctionalities have emerged from a dynamic and indeterministic system, these problems would be circumvented. This implies that the dysfunctionalities of language do not necessarily jeopardize the account of language evolution by virtue of natural selection.

The algorithm of Turkel's simulation is quite straightforward and mostly intuitive. Most parts of the algorithm are quantitatively the same as Hinton & Nowlan; initially 200 agents are prepared. The ratio of [0]:[1]:[?] is different in his four different configurations of simulations –2:2:8 (High-plasticity), 4:4:4 (Equal ratio), 3:3:6

(Original), and 6:6:0 (No-plasticity), respectively. Distribution of these genes in an individual agent is randomly decided initially. In the initial population, generally there is no case where two agents hold the same genotype. The reproduction process includes one-point crossover with 20% probability. No mutation is included.

3.2.2 Kirby & Hurford

While Turkel (2002) shows the power of natural selection in evolution of the LAD, Kirby & Hurford (1997) study the evolution of the LAD from a slightly different point of view. More precisely, they show that natural selection is incapable of shaping the LAD even if the mechanism enables a population to gain better fitness eventually. Instead, they state that language evolution has occurred hand in hand with historical changes in languages. The basic mechanism of their model is similar to Hinton & Nowlan and Turkel. In their model, all possible grammars are coded by eight bit strings. The LAD is also coded as an eight ternary digit array of genes consisting of [0], [1], and [?] alleles. A population is spatially organized based on the organization, both learning and communication take place (Figure 3.1).

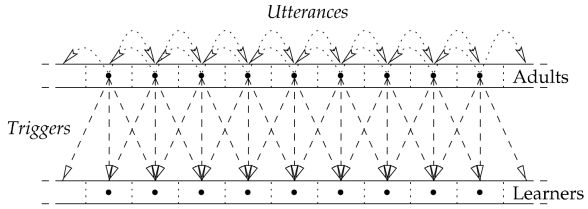

Figure 3.1: The Spatial Organization of Kirby & Hurford

To acquire a target grammar, the LAD changes its [?] alleles to [0], or [1] according to input data. Other [0] and [1] alleles are thought of principles; thus no learning takes place on these alleles. This period is considered to be the learning period. While in Turkel, the learning and the communication processes are in a way 'merged', in Kirby & Hurford, the learning process is completely independent of the communication process. Also, in Turkel's model, agents 'acquire' their grammars from their partners, and such information is never passed onto the next generation, while in Kirby & Hurford, learners receive their inputs from their parents' generation. In other words, Turkel implements a type of horizontal transmission in his language evolution model, while Kirby & Hurford adopt a mode of vertical transmission. It is obvious that regarding language acquisition, the vertical transmission is more important than the horizontal transmission. In this sense, Kirby & Hurford's model is more plausible.

During the learning period, each agent is provided with 200 linguistic inputs. All inputs are randomly derived from three adults' grammars. These adults are

neighbors of the learner; one of them is the previous occupant of the position of the current learner and the other two are direct neighbors of the adult. Any one of the inputs includes only a bit of a grammar. All other bit information in the other positions is masked (*e.g.*, <...*.*.[1].*.*...>). By changing [?] alleles, learners try to parse the utterances so that they form their own grammars. At the beginning of the learning period, a learner expresses a grammar based on her genotype; all [?] alleles randomly express either [0] or [1] in her grammar. Then an input is compared with the grammar, and if it is not accepted (*i.e.*, the values are different), and its corresponding allele in the genotype is plastic, the grammatical value of the position is modified. This learning algorithm is based on Wexler & Culicover's (1980) "*Trigger Learning Algorithm*" (**TLA**) with some modification (the modified TLA: *m*TLA). Effectively, this masking system enables learners to converge on slightly different grammars from the adults grammar. Also, the spatial organization produces some dialect effect.

Then, based on its grammar, each agent attempts to communicate with another agent. Their fitness is calculated based on communicability of agents; two agents who are neighbors of each other compare their randomly selected one bit information of their grammars. The method is basically the same as learning. In contrast to Turkel, therefore, matching of two grammars (adult's and learner's) itself does not affect the fitness value directly.

These mechanisms bring language change into a glossogenetic span (Hurford 1990) since it does not require perfect learning nor communication. It is possible that, even if grammars are different between an adult and a learner, the learner may parse the adult's inputs. Thus one generation's language is not guaranteed to pass through to the next generation with 100% accuracy. This represents a language change through the bottleneck effect. Figure 3.2 shows the overview of the model[5].

Kirby & Hurford also introduce an interesting trick in both learning and communication. In both processes, the first four bits of the arrays are stochastically biased so that [1] is encouraged to be filled in the positions. In the learning period, when [0] is received as an input while the corresponding grammatical information is [1] and its allele is [?] (so with the input, learning can modify the grammatical information so that the input can be accepted), with a minor probability, [1] is retained in the grammar. Consequently, the frequency of [1] in the first four bits of grammar increases. They conceive of this as a parsing bias. Similarly, in the communication period, with some probability, the number of [1] allele in the first four loci of a grammar affects fitness. This can be thought of a communication bias.

[5]Both Figure 3.1 and Figure 3.2 are taken from (Kirby & Hurford 1997).

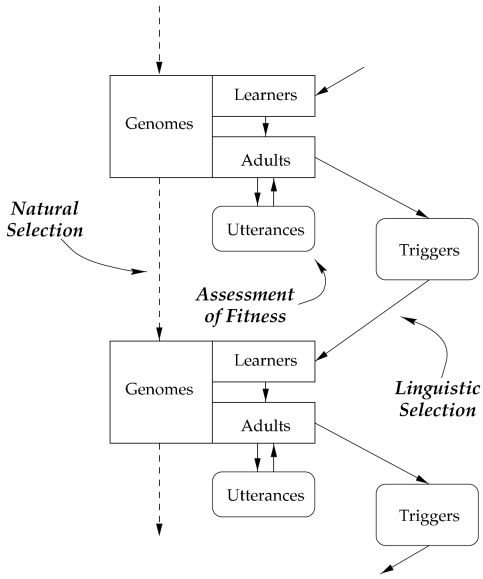

Figure 3.2: The Overview of Kirby & Hurford

The results show that the canalizing effect reliably takes place in both configurations (*i.e.*, both biased and non-biased configurations). Especially when the bias is introduced, the first four alleles in most genotypes in later genepools are canalized. This suggests that, through cultural evolution, selective pressure on language acquisition is properly transfered into the phylogenetic pressure. This view is assured from the result that when the bias in acquisition is removed, but the same bias in communication is retained, no biased canalization is observed. From this, Kirby & Hurford made the following two important statements: "*From initially random initial conditions, linguistic selection leads to a glossogenetic adaptation of the languages in the arena of use*", and "*This glossogenetic adaptation enables the phylogenetic adaptation of the LADs in the population through the Baldwin Effect. Over time, some of the regularities in the linguistic input become nativised*" (Kirby & Hurford 1997). Finally, replication of this study will be given in Chapter 9.

3.2.3 Yamauchi

Both Turkel's and Kirby & Hurford's simulation models successfully demonstrate that the Baldwin effect may take place in the domain of language evolution. Both simulations show that the number of plastic alleles decrease as generation goes. Especially, in Kirby & Hurford, grammatical information which is initially transmitted by learning is quickly taken over to genetic inheritance.

Given the fact that the simplicity of Hinton & Nowlan's model enables it to be the major achievement in the study of the Baldwin effect, the results of both Turkel and Kirby & Hurford, whose architectural designs involve almost minimum

modifications from Hinton & Nowlan, should be taken seriously. However, it is equally true that the degree of abstractions of the models is non-trivial. This is particularly considerable since the models are meant to be the models of language evolution, but not the models of some more general cognitive activity like Hinton & Nowlan modeled.

One of the aspects of language that seriously suffers from these abstractions would be the genetic representation of LAD. In both Turkel and Hinton & Nowlan, each principle/parameter has its corresponding allele in a genotype. However, somewhat apparently, this abstraction is highly unrealistic; as the most complex cognitive ability, the possibility that the 'unit' of linguistic knowledge (*i.e.*, such as principle/parameter) directly corresponds to the unit of genetic information, namely genes should be diminishingly small. Rather, at best, a set of genes may contribute to express one of such units. This type of indirect encode mechanism of phenotypic information brings G-P decorrelation. Consequently, as Mayley (1996*b*) demonstrates, the canalization process may well be blocked.

To examine this possibility, Yamauchi (1999, 2001) conduct a simulation based on Turkel's language evolution model. G-P decorrelation is implemented by introducing epistasis in to the model. Although a detailed description of the simulation model and the *NK*-landscape mechanism is provided in Chapter 9, the result of the simulation evidently shows that the canalizing effect is impeded as the degree of epistasis increases. The same result is also obtained when the same epistatic mechanism is introduced in the model of Kirby & Hurford. This simulation is replicated in Chapter 9.

3.2.4 Briscoe

Since 1997, the computational linguist and language evolutionist Edward Briscoe (1997, 1998, 1999*a*, 1999*b*, 2000*a*, 2000*b*, 2002*b*, 2003) has established a highly sophisticated simulation framework of language evolution. Based on a GA, he builds simulation models whose linguistic base is the framework of "*Categorical Grammar*" (**CG**). CG is a school of linguistic formalism which is equipped with highly enriched lexical inventories, in lieu of relying on autonomous grammatical rules. It has been claimed that this way of describing linguistic behavior is more syntagmatically natural than the orthodox approaches. At least several sub-branches exist in this framework, and each one of them has a different view especially regarding the contents of lexical inventories and basic bounding schemes, but their overall views are fundamentally the same. In this section, we review some results of his simulations. Although there are some different types of simulations, as the overall structures are very similar, here we concentrate on describing two simulation models

(Briscoe 2000*b*, Briscoe 2002*a*). As the models are highly complicated, this section only deals with key concepts in the models and reviews the results.

The representation of a language is based on the framework of CG. One of the features of the framework is found in its enriched concept of syntactic rules. Different from the transformational generative grammar where long-distance dependencies are described by syntactic rules that connect constituents more or less directly, in the CG family, lexical items are locally bound by relatively small sets of syntactic rules. Therefore, the role of lexical items is relatively high in the framework; lexical items are recognized as types of syntactic categories. Other categories also include non-lexical items. A syntactic rule is applied to combine an argument category (*i.e.*, a lexical item) and a functor category (*i.e.*, a non-lexical item). This creates a derived category. Through this unification process, lexical items (*i.e.*, elemental categories) and derived categories are combined to make more complex categories. Such unifications are effected by some small sets of operators. Operators are syntactic rules that tie up different categories. The unification process continues until it reaches the most fundamental linguistic structure which will be fed into the semantic process delivering LF (logical form). In this sense, no dynamic transformation which is manifested from a lower syntactic structure exists. Instead, rules are directly applied on adjacent neighbor lexical items and phrasal categories. "*Generalized Categorical Grammar*" (**GCG**) is a derivation of such a framework.

Notably, there is some form of hierarchy in these syntactic rules. While some rules define general directions of a functor's application, others may override such rules in more local cases. With a relatively small set of these rules and lexical items, GCG can successfully describe different word orders among languages. Thus, the representation of language is far more realistic than Kirby & Hurford, for example. Indeed, given this complex representation of syntactic rules, there are classes of languages; some are subsets of other more 'general' languages, and even there are even 'impossible' grammars.

A language is a set of strings which can be analyzed with a parser based on its associated grammar. The same string can be analyzed by different grammars. Some may yield the same derivation, and others may end up with different analyses. Grammars are also capable of creating sentences. Importantly, in contrast to Turkel, and Kirby & Hurford, the complex representation of language can allow it to be a complex adaptive system; not only agents but languages can also adapt to those agents so that they are easily learnt. The arena of use is defined by a mixture of various sentences.

In the models, an LAD is represented by principles and parameters called "*p-settings*". Each principle/parameter represents a syntactic category or the property

of a functor. There are twenty such *p-settings* in one genotype that configure different types of languages. In Briscoe (2000*b*), all *p-settings* take ternary values. All syntactic categories and rules have binary allelic values. Therefore a parameter can be either one of such values or a completely unset, neutral value. This means that the notion of parameters is slightly different from other simulation models; principles are those *p-settings* which cannot be updated at all, while non-neutral, valued parameters are resettable upon learning. Such valued parameters are called 'default' while non-valued parameters are called 'unset'.

Unset parameters in this model are closer to those parameters used in Turkel and Kirby & Hurford. On the other hand, the idea of default parameters is somewhat similar to the *marked-unmarked* distinction in the study of language acquisition. By a given learning algorithm based on TLA, such values are updatable for given periods (depending on a configuration of the simulation) upon inputs. With one trigger, only n parameters are updatable. Therefore, if no input affects a given default parameter, such a parameter expresses the default value of an associated grammar. Parameters are updated upon parse failures. In Briscoe (2000*b*), he prepares two different types of update algorithms in terms of the number of updatable parameters. When n=1, the algorithm is incremental so that even the updated grammar cannot fully parse a given sentence yet, it improves its parsability, the agent retains the new settings. On the other hand, when the number is n=4, four randomly chosen parameter settings are retained if the updated grammar can fully parse the sentence. The learning algorithm is also partially ordered as some *p-settings* represent more general/fundamental properties of syntactic information.

In Briscoe (2002*a*), such ternary *p-setting* values are replaced with probabilities. Such probabilities correspond to likelihood for a certain value to be set in the principle/parameter. With a Bayesian learning algorithm, such probabilities are updated, instead of adopting TLA. The probability assigned to a principle or a parameter determines what type of specification a particular operator takes.

In contrast to other simulations described here, the model also has a complex population structure. First it has migration. The spatially distributed agents sometimes migrate to different locations, which triggers a shuffling at phenotypic level. Also, at any given time in a run, four different generations are in one population. Moreover, learning and communication simultaneously take place within a single population; agents have 10 time-steps of lifetime and for the first four steps, agents can learn from adults. This is an implementation of the critical period. On the other hand, communication takes place in any period. All time-steps are increased equally across the four different generations in a given population. In Briscoe (2000*b*), agents may 'die' between 5-10 in proportion to their fitness values.

The reproduction process takes place between 5-10. Those who reach step 10 are obligatorily removed from the population. This complex population system creates a dynamism in language evolution. In Briscoe (2002*a*), all agents can fulfill their lifetime.

The fitness function is also complicated. There are seven different factors involved in the function; costs are involved in sentence production, production of subset language, parsing, parsing failure, memory cost, and parameter update. The benefits are given by interaction success, and successful parameter update. Therefore, not only communicative successes, but also learnability which is calculated by successful parameter update divided by parameter update cost (times the inverse of maximum updatable parameters, *i.e.*, n), and expressivity (communication by using subset languages are penalized) are involved.

Migrations introduce language changes in the population. Typically, a population converges to a small set of languages and no notable move takes place any further. At any given time, one third of a population is replaced with the same number of adults who have a different language. However, their genetic component is the same as the dominant genotype at the time. Therefore, a migration brings linguistic diversity, yet the current genetic diversity is retained. If the frequency of migrations is high, it brings rapid language changes in the population. Regarding the Baldwin effect, this gives an interesting insight as the categorization effect considers. Under such a fast changing environment, no significant canalization process, but some degree of the categorization process take place.

With these models, Briscoe has conducted various simulations. From the point of view of the Baldwin effect, his simulations are interesting as they properly implement linguistic selection since languages in the simulations are complex systems which dynamically adapt to language learners. Various factors coded in the fitness function either directly or indirectly make complex tensions in the axes of learnability, expressivity, communicability, and parsability.

In the simulations where no linguistic migration is implemented, and thus where the linguistic environment is rather constant, typically unset parameters are quickly driven away while default parameters show a sharp increase. On the other hand, principles keep constant from the beginning. As unset parameters are the least informative, increase of default parameters is a clear case of the canalizing effect. He also reports that evolution of a population is always in the direction of initial *p-settings* that increase the learnability of the dominant language in a given environment (Briscoe 2000*b*). Therefore, while unset parameters diminish, those increasing default parameters keep the grammatical values of the dominant language in the environment. In the simulation, the number of updatable parameters also evolves;

with the same mutation probability of the *p-settings*, the number can mutate ±1. The mean number of updatable parameters per trigger also decreases until the population converges to have a value of 2 or 3. This consequently contributes to increase fitness as a cost is incurred in every parameter update.

When the migration process is introduced, rapid linguistic changes take place. As a result, replacements from unset parameters to default parameters are less likely, and replacement from unset parameters to absolute principles become more common. However, in an absolute sense, overall evolution of canalization is slow and less complete.

Given the result of Yamauchi (1999, 2001), Briscoe (2002*a*) conducts simulations in which a type of G-P decorrelation scheme is introduced; a mutation affects more than one *p-setting*. Using much the same idea of the *n p-setting* update scheme, a mutation can modify n numbers of *p-settings* during the reproduction process. The more *p-settings* are modified by a single mutation, the less correlation exists between a genotype and its corresponding phenotype. Since this consequently brings linguistic changes in the population, in half of the runs, no migration is considered in the model.

Generally, decorrelations bring less expressive subset languages to the population. This is due to the high rate of language changes caused by effective mutations. Timings of linguistic changes seem to correlate with appearances of mislearning agents (in a given language environment). However, new languages caused by such linguistic changes create chances for other agents who have similar genotypes to the mislearners to enjoy acquiring the new language. Thus, decorrelation creates a new linguistic selection for more learnable languages.

Briscoe reports that the mean percentage of mislearning agents who fail to acquire a full grammar or grammars is under 1% for low decorrelation, 4.5% for intermediate, and 24% for the highest. Under the highest decorrelation configuration, every run shows that the population converges to a minimum subset language.

Interestingly enough, the number of default parameters which evolve from unset parameters actually *increases* by roughly 5% under different configurations. However, Briscoe reports that the standard deviation from the mean generally increases.

While this increase of default parameters would be counted as a case of the canalizing effect, if the timings of such shifts and linguistic change are compared, a slightly different view appears; often mutations are not *assimilative* (*i.e.*, adapt to the current environment). Rather, such mutations are the *causes* of linguistic changes. Once such changes prevail, these *preemptive* mutations become indistinguishable from assimilative ones. Briscoe reports that for low rates of decorrelation,

roughly 20 % of mutations which successfully prevail in the population are preemptive (*i.e.*, 80% are assimilative). For intermediate, and the highest, 45% and 99% of fixated mutations become preemptive, respectively.

Basically this is consistent with Yamauchi. However, given the rich linguistic representation, a more interesting insight is available from the results. First, decorrelation not only disturbs the canalization process, but also keeps the population in a suboptimal state regarding linguistic expressivity. This is because subset languages are more benign to learners whose *p-settings* are incompatible with the current languages (due to decorrelated mutations). Recall that in the canonically correlated model, *p-settings* are assimilated to dominant languages in a population, and subsequently this restricts future linguistic diversity (unless the migration process is introduced). This coevolution process is also found in the decorrelated situation. However, due to the effective (deleterious) mutations, the coevolution process under this circumstance works in a different way; mutations inhibit learnability in the linguistic environment at a given time. As a result, rapid linguistic changes are triggered. Some mutated *p-settings* which did not trigger such linguistic changes could become adaptive under a new linguistic environment brought by the preemptive mutations. Those assimilative *p-settings* are in a sense properly canalized. Therefore, mutations serve for both creating a new linguistic environment and assimilating to the extant environment.

Chapter 4

Reconsideration of The Baldwin Effect

4.1 Baldwin Skepticism

4.1.1 Lack of Explananda

In the previous chapters, we argued some basic points of the Baldwin effect from both biological and computational points of view. Given these understandings, we have also reviewed some linguistic investigations adopting the concept. As in the example of Baldwin himself, many scholars who are concerned with higher order psychological abilities such as learning, conscious mind, or morality are generally fond of the concept. This is noticeably true after Hinton & Nowlan; these interests have been the main thrust of investigations of the conceptual possibility. However, more than ten years of grossly positive attitudes have recently provoked some backlash towards the effect. This attitude is sometimes called "Baldwin Skepticism" as opposed to "Baldwin Boosterism" (Godfrey-Smith 2003). Interestingly one of the earliest skeptics of the Baldwin effect comes from Simpson (1953) himself; while he introduced the Baldwin effect in modern biology, his attitude was rather dismissive. His main point was that although the Baldwin effect itself is not a theoretically fallacious concept, its application to modern evolutionary biology is unnecessary. Simpson clearly expressed his attitude in the following statements:

> ... each process necessary for the Baldwin effect does factually occur. There is no reason to doubt that they could occur together, in the stated sequence, and so produce the Baldwin effect. There is even some probability that they must have produced that effect sometimes. Nevertheless two points remain decidedly questionable: whether the Baldwin effect does in fact explain particular instances of evolutionary change, and the extent to which this effect has been involved in evolution or can explain the general phenomenon of adaptation.
> (Simpson 1953, p. 103 in the reprint)

More precisely, Simpson considered that there is no need to invoke any special evolutionary mechanism to explain currently known evolutionary phenomena:

> The Baldwin effect is fully plausible under current theories of evolution. Yet a review of supposed examples and of pertinent experiments reveals no instance in which it indubitably occurred, no observations explicable only in this way, and few that seem better explained in this way than in some other.
> (Simpson 1953, p. 106 in the reprint)

and finally:

> The Baldwin effect is both possible and probable but assignment to it of that role in evolutionary theory seems to me fallacious.
> (Simpson 1953, p. 106 in the reprint)

From these statements, especially from the last, it is clear that Simpson casted a strong doubt on its contribution to the explanatory power of Darwinian evolutionary theories. The most prominent Baldwin skepticisms in modern times share this point. Downes (2003) nicely describes the situation with a comparison to Gould & Eldredge's (1977) theory of punctuated equilibrium. Immediately after the introduction of the theory, it was put through serious debates and examinations with available data. This is because the theory was expected to solve an apparent contradiction between empirical data derived from fossil records and purported accounts provided by normal Darwinian gradualism; a number of fossil records indicate that evolution shows rapid and radical changes followed by long stagnations. The theory of punctuated equilibrium provides an attractive account for the data.

Downes' criticism is that, on the other hand, the Baldwin effect lacks proper empirical objects to be explained. Dennett (1991, 1995) is also targeted in his criticism. Downes discusses the fact that although Dennett emphatically promotes the Baldwin effect as an evolutionary mechanism which enables evolution of consciousness, he fails to provide even a single concrete example in his arguments. While Dennett (2003) attempts to avert Downes' claim, so far he has not provided sufficient empirical evidence, which is exactly what Downes criticizes. If this is indeed the case, the concept is redundant in evolutionary biology.

However, this type of criticism may not be fundamentally so critical. It would be highly plausible that possible candidates of explananda are to be found in the future; generally, detailed evolutionary studies have been conducted with either relatively simple organisms, or in comparatively monotonic interactions with environmental factors. On the contrary, what proponents of the Baldwin effect are interested in most are typically evolutionary processes of complex cognitive systems or more dynamic interactions between organisms and environments through plastic behavior, such as evolution of traditions or cultural inheritance where extragenetic inheritance

also takes place. Because of practical constraints of such studies (*e.g.*, feasibility of experiments, long lifespan of higher order organisms, or lack of methodologies, etc.), our understanding of the evolution of higher order cognitive capacities is still highly limited. However recently the basic principles of evolutionary developmental biology have been applied to psychology and created a new enterprise in psychology (evolutionary psychology). It shows a dramatic growth of interest from surrounding fields (Griffiths 2001). In this regard, the situation is radically different from the period when Simpson or Waddington conceived rather primitive interactions between phenotypic plasticity and genic expressions. Although it becomes difficult to find clear causal relationships in higher order behavior, especially mechanisms of canalization of learnt behavior, it is highly plausible that nothing but the Baldwin effect could provide a suitable account for given data. It is even more conceivable that the evolution of language contains some explananda for which the Baldwin effect provides a proper explanation. After all, among evolutions of such psychological abilities, the evolution of language is certainly at the height of complexity.

4.1.2 Methodologies

The second type of skepticism revolves around a methodology often used in the studies of the Baldwin effect, namely the constructive computational approach. The proponents of the Baldwin effect often adopt the computational approach to investigate the concept. This is mainly because interactions between learning and evolution are often highly complex and dynamic. Subsequently, their causalities are often beyond our understanding. These are the fields where the constructive computational approach has proven to be a powerful tool; as we have seen in previous chapters, the approach enables us to investigate such systems with a small number of factors and clear causalities.

The core parts of the approach are modelings and simulations. However, these methodologies (*i.e.*, modeling and simulation) have a non-trivial downside. The constructive approach carries out its research procedure in a rather different way from the conventional reductionistic approach. Among many differences between these two approaches, the most significant difference is incomplete understandings of research objects. As stated above, this is largely due to their complexity and dynamics (or time constraints). With these limited understandings, we hypothesize models with necessary idealizations and abstractions. Thus keeping a given object largely as a black box, we extract known constituents from the object which we consider important. Based on the constituents, then a model is constructed. The model is an analogy of the object. With the model, simulations are conducted. Since the simulations are interactions of the extracted constituents, the behavior of such

interactions can be interpreted as "*as-if*" miniatures of the objects. As the models are idealized and abstracted extractions of the factors in the real objects, causalities in the simulated behavior are far easier to understand than the real objects. By evaluating the results, models are refined; and consequently understandings of the objects themselves are improved.

It is often the case that abstractions of real objects are non-trivial. The analogies between the model and the real world produced by these idealizations and abstractions often confuse some researchers who are unfamiliar with the approach. For instance, in the same paper, Downes (2003) criticizes the results of Hinton & Nowlan. He claims that the results of the model cannot be taken as evidence that the Baldwin effect adds a new explanatory power to the standard Darwinian evolutionary theory. His main point regarding the model is Maynard Smith's interpretation of the result. Recall that Maynard Smith interpreted the result and claimed that learning speeds the pace of evolution (*i.e.*, the expediting effect, Maynard Smith 1987). Downes rejects this claim on the basis of the abundance of such examples in nature:

> So in a computer model with many idealizations, we can simulate a huge increase in the speed of evolution. But does an increase in the rate of evolution force us to propose a new evolutionary mechanism? The answer to the question is no. There are many examples of rapid evolution. If we rule out asexual cases, we still have numerous examples. (Downes 2003, p. 19)

However, this argument should be read the other way around. Phenomena are the explananda, and mechanisms are the explanans. In nature, there are many instances that exhibit rapid evolution (*i.e.*, the explananda). Possible mechanisms for this may be pluralistic. If a simulation demonstrates that the interaction of learning and evolution is possibly one of the mechanisms, then it *per se* provides a reason for an investigation of the particular mechanism. An analogy may clear up the point; there are a large number of species that can fly (or at least glide). Obviously, more than one architectural mechanism is responsible for the ability (*e.g.*, birds' feather wings, pterygote's veined wings, or flying fish's large pectoral fins). Of course all of these follow the basic rules of aerodynamics (the causality level shared by all flying species), but their physiological structures are quite different and consequently methods of flying and gliding are also different (the level which we are interested in). For instance, flying patterns of butterflies are distinctive from any other species. Then, investigations of these mechanisms should be, at least in the practical level, separated. Much in the same vein, if a given trait is thought of exhibiting a rapid evolutionary history, mechanisms of expediting effects

in evolution should be individually treated and investigated. It is true that asexual species are different from sexual species regarding the speed of evolution. However, the mechanism involved in asexual selection does not account for how learning accelerates evolution. Dealing with the expediting effects in evolution in a single lump is hazardous.

4.1.3 Conceptual Ambiguity of the Baldwin Effect

The third point of the criticisms is the conceptual ambiguity of the Baldwin effect. One may note that the usage of the word "effect" in the Baldwin effect is quite different from a normal sense. Typically, the term is associated with observable, physical phenomena, such as "the Doppler effect". Such phenomena are theory-*neutral*; they are objective facts and their existences are independent of observers' theoretical principles. On the other hand, it is somewhat difficult to conceive of the Baldwin effect as a real 'effect' in nature. First of all, as in the critiques, it still seems to lack a real instance in nature. Moreover, the concept is multi-faceted, and most properly understood as a collective concept; as we have observed, the Baldwin effect includes at least the expediting and canalizing effects. Both effects can independently take place. Sasaki & Tokoro show that even these effects have complex properties. Avital & Jablonka propose further possible effects especially in the canalizing effect. Depew points out:

> ... these conceptual differences are so great that the Baldwin effect cannot be said to stably refer to a single process, either empirically or conceptually, as it migrates from the theoretical field in which Baldwin himself, for example, placed it to the quite different theoretical background that Simpson assumed.
> (Depew 2000, p. 9)

However, the Baldwin effect should be considered as a type of "*umbrella term*" –a term used to cover a broader category of phenomena rather than referring a specific phenomenon. One such concept in evolution is natural selection (Corning 1998); the term "natural selection" encompasses any processes causing the differential reproductive or survival successes regarding genes, genotypes, populations, or species in functionally significant ways (as opposed to, say, random, stochastic processes)[1]. In this regard, umbrella terms would be, in general, inherently ambiguous. On the other hand, such terms are able to encompasses a range of phenomena under a single concept. As long as the identification of a term is correct (as in the term

[1]Often, changes of gene frequencies in a genepool are identified as the case of natural selection; this is somewhat inadequate in the fields of non-population genetics.

natural selection, for example), the umbrella term enables us to concentrate on investigating a higher-order regularity by ignoring differences in individual instances and highlighting similarities.

Given this, the problem of the definition of the Baldwin effect becomes clear. While the term bundles together a collection of evolutionary phenomena, it lacks the all-agreed common handle; the core concept that encompasses all instances covered by the term. In the example of natural selection, the handle is the differential reproductive success attributing to functionally significant mechanisms. What is the handle of the Baldwin effect? It would be agreed by all that the core constituents are something produced by interactions between learning and evolution.

A hint of this question may be found in the term "*synergy*" of learning and evolution used by Turney *et al.* (1996, 1996*b*). According to Peter Corning (1983), synergy is a pan-disciplinary concept and is defined as "*the effects produced by wholes are different from what the parts can produce alone*" (Corning 1998, p. 135). By introducing this concept, together with viewing the Baldwin effect as an umbrella concept, there is a necessary shift in focus from individual instances to the cooperative behavior of constituents that produces various instances of effects as its results. Under this scheme, the Baldwin effect can be recaptured as the synergistic effect of learning and evolution itself. By doing this, the Baldwin effect is broadly defined as a collective, multi-faceted concept of the synergy of learning and evolution; under the same handle (*i.e.*, the synergistic interactions of learning and evolution), it covers different types of effects (*i.e.*, a proper umbrella term). It is conceivable that the Baldwin effect is a synergistic effect of learning and evolutionary search. Both the expediting effect and the canalizing effect are the products of their synergistic interactions because a mere combination of learning and evolutionary search may be insufficient to produce such effects. Turney (1996*b*) criticizes that researchers too easily conceive any combination of learning and evolutionary search *is* the Baldwin effect. He concerns that by equating the coexistence of learning and evolutionary search in a population with the Baldwin effect, researchers may neglect important aspects of the Baldwin effect. Crucially, it has been experimentally demonstrated that the Baldwin effect is not equivalent to simple combinations of learning and evolution. In other words, combinations of learning and evolution do not necessarily produce the Baldwin effect as Mayley demonstrates (Mayley 1996*a*, Mayley 1996*b*, Mayley 1997).

This beams a light into a proper investigative avenue for the concept. Mayley's series of work, for example, can be now clearly understood as studies of the necessary conditions of the synergy; when, and under what conditions, do learning and evolution start to cooperate; what are the consequences? Especially in his study of

G-P decorrelation (Mayley 1996*b*), Mayley convincingly demonstrates that sensitive conditions of search spaces both in learning and evolution are required for the synergistic behavior to be yielded. Also, by definition, possible mechanisms of the synergy are pluralistic.

Therefore, careful examinations of purported models of the Baldwin effect reveal those models stipulating different types of mechanism in terms of interactions of learning and evolution, and canalization. In the section, we examine possible mechanisms of the Baldwin effect.

4.2 The Mechanisms

Without recognition of the Baldwin effect as an umbrella concept in evolutionary study, actual mechanisms of the effect have not been considered seriously. For example, computational studies have revealed that the Baldwin effect is indeed a feasible concept, as shown in Section 2.8. However, most studies of such models focus on the results, but scarcely examine the actual mechanisms adopted in the models. Like Hinton & Nowlan, if the primary purpose of a model is just to present the feasibility of the concept in general (*i.e.*, not for a specific instance in evolution), a mechanism of the synergy could be highly idealized. However, if some evolutionary phenomenon which might contain the synergistic interactions of learning and evolution were investigated, a model containing a plausible mechanism of the synergy would have to be carefully designed.

So, what are the possible mechanisms of the Baldwin effect? As more detailed questions are asked, it becomes evident that the explanation of the causality is not exhaustively expressed. For example, one may ask what type of interaction works between evolution and learning? Is it a direct relationship? Is it indirect? Does learning only refer to post-natal adaptation? It is often the case that researchers ignore (or simply do not realize) these points, and argue possible impacts of the effect as a whole. The pros and cons of the Baldwin effect depend on which mechanism is referred to as the Baldwin effect. Thus, boosterism and skepticism towards the Baldwin effect are quite susceptible to a type of concept we bear in mind. For example, there are two types of selections in evolution; natural and sexual selection (and possibly the third type of selection, namely artificial selection). Although the fundamental concept of these two selections is the same, they are better classified separately as the mechanisms and working situations are quite different. A similar argument can be made for the Baldwin effect. While Simpson (1953) 'formalized' the effect, the formalization merely states what type of a chain of effects can be called "the Baldwin effect". It does not state what type of cause produces the effect. Some mechanisms can be candidates for the cause of the effect. In the following

sections, three different mechanisms of the Baldwin effect are discussed, based on the study of Godfrey-Smith (2003).

4.2.1 Baldwin's Breathing Space Model

Baldwin believed that if an organism can 'accommodate' to its surrounding environment, it may survive longer[2]. Subsequently, such an organism can obtain more breeding opportunity and thus, the population itself is kept alive by virtue of such individuals. Eventually, among the members of the population some individuals' plastic behavior would be replaced by hereditary behavior –organic selection. He defines this selection as follows:

> *Organic Selection*: The process of individual accommodation considered as keeping organisms alive, and so, by also securing the accumulation of variations, determining evolution in subsequent generations.
> (Baldwin 1902)

This is the reason that Baldwin called accommodation "a new factor in evolution". This new factor is directly comparable to the concept of adaptive (*i.e.*, hereditary) behavior, since both accommodation and adaptation contribute to an organism's struggle for survival directly, and given this, its breeding chance can increase. Eventually, natural selection favors those individuals equipped with the adapted behavior innately.

This type of mechanism for the Baldwin effect is sometimes called a "*Breathing Space*" type mechanism (Godfrey-Smith 2003); learning keeps the population alive long enough to provide time for breeding (*i.e.*, breathing space). However, this type of Baldwinian mechanism would be difficult to defend, as it requires some strong assumptions. First, the environment must be harsh enough so that learners have a definitive advantage over nonlearners in terms of survival and reproduction. Otherwise, the population might not make the transition to the second Stage in Simpson's formalization of the Baldwin effect because other non-learning individuals pull the population back to Stage 1. Second, even in a severe environment, such learners, whose number is initially presumably small, have to save the whole population. This is also true when the population moves from Stage 2 to Stage 3. This assumption, however, is somewhat strange. The reason the population successfully moves from Stage 1 to Stage 2 is that learning is truly advantageous in a harsh environment. The question is, then, why do such a small number of individuals, enjoying innate equivalent or similar behavior, outperform the learners

[2]His usage of the term "accommodate" is equivalent to 'acquired behavior' in the current study of evolution. Thus the difference between normal adaptation and accommodation is hereditary or non-hereditary, respectively.

in Stage 2, so that they can indeed push the population to Stage 3? In Stage 1, non-learning individuals die out before they can breed. The population in Stage 2 is sufficiently fit for the environment. Hence, the environment is no longer too harsh for the population; selective pressure is now greatly weakened. Consequently, it is hard to imagine why such hereditary behavior prevails against learnt behavior in the population.

For Baldwin, learning is evolutionarily advantageous simply because it 'keeps the population alive' the population; his original concept apparently did not incorporate the concept of a heredity element. Concepts of developmental or population genetics were simply not available in his era. In the next section, we will review Waddington's mechanism of the Baldwin effect which was formed in the period of the Modern Evolutionary Synthesis.

4.2.2 Waddington's G-P Correlation Model

Waddington's mechanism of genetic assimilation (and more specifically, canalization) is undoubtedly the most popular mechanism employed in Baldwinian explanations of evolutionary processes. While Baldwin's original theory lacked the basics of genetics, Waddington's genetic assimilation is theorized within the framework of the Modern Synthesis; the formulation of the theory is thought of genetically feasible. He experimentally proved that a genetic assimilation indeed takes place, while it has yet to be confirmed in nature.

As we have discussed, his original theory of genetic assimilation is often mixed up with the current mutation-centric view in computer simulations. However, his own mechanism is a more static one. For Waddington, a given genepool retains enough genetic variance so that the target genotype easily surfaces through comparatively small numbers of sexual recombinations (or else the original genepool already holds the particular genotype in a small proportion). Therefore, in contrast to Baldwin's original formulation, Waddington's mechanism leaves the possibility that a population in Stage 1 makes the transition to Stage 3 without passing through Stage 2 (*e.g.*, in the case of Waddington's experiment of *Drosophila melanogaster*, the normal wing population directly shifts to the innately equipped population of such wings). However, as the advantageous nature of learning over evolutionary search is the very heart of the Baldwin effect, such a case should be highly unlikely in the examples we consider. Therefore, in Waddington's mechanism, it is as if learning acted to mediate these two separated stages (*i.e.*, Stage 1 and Stage 3).

However, for genetic assimilation to take place, two crucial conditions have to be met. The first condition is the existence of phenocopy. Phenocopy is, as described in Chapter 2, an environmentally induced trait that closely resemblances a given

heritable trait. At a glance, this first condition looks somewhat self-evident; in Waddington's formulation of genetic assimilation, if a selected learnt trait does not resemble any genetically heritable traits, it is impossible that selection for such a learnt trait ultimately leads to canalization. This condition is also required in any other mechanisms of the Baldwin effect.

Secondly, even more importantly, such a learnt trait has to be genetically closer to the innately predisposed trait than other non-learnable, no-innately-predisposed individuals (*i.e.*, the population in Stage 1). In other words, a given innately predisposed trait and its phenocopy have to be genetically related; such phenocopies are not only functionally close to that of innately predisposed ones, but also they are genetically closer than those of non-learnt phenotypes.

Here is an example. Suppose there are three cities in a certain region, called A, B, and C. Those cities are roughly on the same line (say, Washington D.C., NYC, and Boston). Starting from A, heading for B also means getting closer to the city C. If, however, those cities are not on the same line, but rather scattered, the situation would be different (say, Washington D.C., NYC, and Chicago). Heading for B from A does not mean C is getting close. Bearing this in mind, consider the same situations in Waddington's mechanism. In Waddington's case, the genotype of the plastic flies is closer to the genotype of the innately equipped flies than that of the original normal flies. In other words, a linear relationship exists among those three genotypes. On the other hand, if the relationship is non-linear, getting close to the genotype of plastic flies from the original normal genotype actually increases the distance from the genotype of innately equipped flies. Then if the population completely shifts to Stage 2, the probability of such innately equipped flies appearing from Stage 2 would be lower than the probability from Stage 1 (this would especially happen when the relationship is linear but the order is B, A, and C).

Let us look at a more concrete example. To explain the efficiency of Waddington's canalization over other normal evolution, Avital & Jablonka (2000) use the following simple genetic model as an example: Suppose $a^1a^1b^1b^1$ is the most popular genotype in a population where a and b are different genes, and the number shows a particular type of allele. This genotype does not allow for plasticity. The frequencies of a^1 and b^1 are 0.9 each. On the contrary, the frequencies of a^2 and b^2 are 0.1. Only the genotype $a^2a^2b^2b^2$, which is innately equipped, is advantageous in the current environment, however, from the predominant genotype $a^1a^1b^1b^1$ in the population, the frequency of an individual whose genotype is $a^2a^2b^2b^2$, would be very low, 10^{-4} (*i.e.*, one in ten thousand).

If, however, any individuals whose genotype includes either the a^2 or b^2 allele can adapt to the environment through learning, since the frequency of such individuals is significantly higher than the $a^2a^2b^2b^2$ type individuals, and they are more adaptive than the $a^1a^1b^1b^1$ type individuals, after a reasonable amount of mating, the $a^2a^2b^2b^2$ type individuals will be obtained. This example contains a more sensitive condition than the example of the three cities; there are adaptive variations of learnable phenotypes. For the canalization process to take place, proximities of genotypes attributed to the learnable phenotypes to the genotype of innately adaptive phenotype have to be the same as the order of adaptivity of the phenotypes. In other words, the same topological proximity between learnable phenotypes and innately adaptive phenotypes also has to be retained in the genotypic space.

Therefore, if a better phenotype's genotype has more Hamming distance than another less adaptive one, it creates a local optima. Consequently, the canalization process will be disrupted. Suppose that $a^1a^2b^2b^2$ is the innately adaptive genotype, while $a^2a^*b^*b^*$ ('*' designates either 1 or 2) is the plastic genotype. From the normal population, the innately adaptive individual is hardly obtained (one in thousand). As discussed in the previous example, reasonable numbers of plastic individuals would be present in the population. Unlike the previous example, however, in this situation canalization will hardly occur; under this circumstance the more $a^2a^*b^*b^*$ becomes dominant, the less likely it is that the $a^1a^2b^2b^2$ will appear, because the $a^1a^2b^2b^2$ genotype is not a subtype of the $a^2a^*b^*b^*$ genotypes.

In the examples listed above, the genotypes for both the innate adaptive trait and the plasticity are thought to be different alleles in the same genes. It is highly natural that such similar but different behavior can be attributed to completely different genes. Or even worse, they are on different chromosomes. For example, if the innate behavior is expressed from the $a^2a^2b^2b^2$, while the plastic behavior is expressed from the $p^1p^2q^1r^1$, completely irrelevant genes occupied in a different (or distant) region in a given genotype. This is a case where genotypes attributed to phenocopies are irrelevant to the genotype attributed to the innate adaptive trait.

This genetic correlation between phenocopy and innately adaptive trait becomes a progressively serious concern to researchers who deal with higher order adaptive behavior; it is this field that most researchers interested in the Baldwin effect want to apply the mechanism. Consider that Waddington's studies are all about physical traits. In all his experimental studies of genetic assimilation, Waddington used *Drosophila melanogaster* to investigate canalization of *physical/physiological* traits. Moreover, such traits are not postnatally 'acquired'. Instead, what Waddington considered in his canalization mechanism mostly takes place during the particularly early stages of individual development. In other words, such reactions are

obtained by breaking STE conditions, and it would not be learning or some sort. Rather, such reactions should be considered a matter of developmental genetics or embryology; typically in this stage, modification of a specific trait is irreversible. This is the reason that Williams (1966) criticized Waddington for failing to distinguish between susceptibility in new environmental conditions and adaptive response to environmental stimuli.

Contrarily, a lot of reflexive behavior, such as the blink reflex, in humans are strongly innately predisposed. Much of such behavior is, however, replaceable by intentional motions which are mostly postnatally acquired. As different regions of the brain control such behavior, it is reasonably assumed that the genes affecting these are quite different. Of course, the relationship between reflex behavior and intentional behavior is highly moot, and has to be argued in a more rigorous way. However, this strikes the most basic point –there is no guarantee that two functionally similar types of behavior, one canalized and the other learnt, can be attributed to two similar genes. Often, for behavioral traits to be completely acquired, a maturation period of some extent is required. In some cases, fixation of traits never happens in an individual's lifetime; a permanent plasticity. Then it would be completely possible that the genotype for innate adaptation is more attainable from the original (*i.e.*, non-learning, non-innately specified) genotype than the genotype for plasticity. This is basically what Mayley (1996*b*), Yamauchi (1999, 2001), and Briscoe (2002*a*) present in their studies of epistatic G-P decorrelation. As presented in Chapter 2 and Chapter 3, Mayley, Yamauchi, and Briscoe demonstrate that the search space of a phenotype and a genotype must be closely correlated so that the Baldwin effect takes place. Otherwise, a population is typically entrenched in an environment with learnt behavior. There are some ways to break G-P correlation and both epistasis and complete separation of innately predisposed and learnt traits' genotypes can effectively introduce decorrelation.

Thus, although Waddington's genetic assimilation (and canalization) model is an attractive, and empirically attested phenomenon which may take a crucial part in the Baldwin effect, his mechanism of canalization requires a sensitive prerequisite, namely strong G-P correlation. This, is an arguable point as it is highly unlikely that such a strong correlation is indeed found in higher order cognitive abilities. This as a whole, casts strong doubt on any Baldwinian accounts of language evolution that utilize somewhat naïve assumptions in terms of G-P correlation. Of course, as an approximation, such an optimistic assumption would be acceptable in some cases, but in most of the literature on language evolution, researchers blindly accept such an unwarranted concept. This point will be computationally examined in Chapter

9. Finally, henceforth Waddington's model is called "the G-P correlation model" as a mnemonic name.

Given these arguments, it is unfortunate that both the G-P correlation model and the term "genetic assimilation" were introduced by Waddington himself. Recall that the term "genetic assimilation" has been confusingly used in the context of the Baldwin effect. Since the notion of the Baldwin effect and the non-mechanistic, but phenomenal notion of genetic assimilation are so tightly bound, a lot of researchers into believing that the concept of genetic assimilation in the second sense (*i.e.*, the mechanical sense) is somehow considered as 'the mechanism' of the Baldwin effect (especially, the canalization process). In other words, the G-P correlation model is believed to be the model that describes the mechanism of the Baldwin effect. In reality, however, the possible mechanisms of the Baldwin effect are pluralistic as we have seen in Baldwin's original description of the Baldwin effect, and some do not necessarily require tight G-P correlations for canalizing existing learnt behavior.

This is where, we think, the greatest danger of current Baldwinian accounts of language evolution suffer; in Chapter 2, we have seen various studies regarding the Baldwin effect. None of the studies properly addresses this point, and upon it a theory is constructed. There are some studies which pay more attention to the mechanical aspect of the Baldwin effect (*i.e.*, Deacon 1997, Avital & Jablonka 2000, Dor & Jablonka 2000, Dor & Jablonka 2001). However, while they are implicitly indicating different types of mechanisms which may circumvent the problem of the G-P correlation model, these researchers seem to fail to realize that the Baldwin effect accepts different types of mechanism and they are indeed proposing different mechanisms.

The notable case is Deacon (1997). In his book, he presents the Baldwin effect in conjunction with the context of the biologist Lewontin's constructivist approach. Although this is a revolutionary view of the Baldwin effect, in hindsight, he does not explicitly promote the idea as a new type of mechanism in the Baldwin effect. In the next section, we examine his view of the Baldwin effect in detail.

4.2.3 Deacon's Niche Construction Model

In *The Symbolic Species*, among a number of other important contributions to the study of language evolution, Deacon (1997) introduces yet another type of Baldwinian mechanism. His new formulation of the Baldwin effect is different from the conventional Baldwin effect (*i.e.*, the G-P correlation based model); indeed all the examples he provides in his explanation are originally introduced in explanations of different types of evolutionary processes by others. Unfortunately, as the book is devoted to language evolution, this new formulation has not been widely recognized.

Even so, the impact is large enough that those who do not pay extra attention to language evolution, but who seriously consider the Baldwin effect, have gradually acknowledged what is written in a few pages out of the over-500-page book. These researchers, mainly philosophers of Darwinian evolution (*e.g.*, Godfrey-Smith 2003), realize that the model of the Baldwin effect is isolable from the argument of language evolution itself (*i.e.*, language evolution is *a* type of explananda in this model), and is applicable to more general evolutionary processes.

However Deacon does not make the causal mechanism of the model clear, as his primary aim in the chapter is not to introduce the model itself, but to describe evolution of linguistic capabilities in brain. Subsequently, his description of the model in the chapter circumvents describing a basic style of the model and leaps into more complex and elaborated instances at a bound. Furthermore, it is also true that although Deacon is clear about the Baldwin effect, what he exemplifies are largely different from what is known as the Baldwin effect (as what Simpson formulated). As a result, it becomes difficult to discern the mechanism itself from the model of the brain-language coevolution theory. However, objectively speaking, Deacon's coevolution theory stands as a new form of the Baldwin effect, regardless of whether or not he himself intends this. This may have affected the slow acceptance of the model in the context of the Baldwin effect. In this section, thus, a more detailed introduction of the model with some background concepts is provided. Note, as the kernel of the model is rooted to more fundamental concepts of evolution, a further explanation for the concepts is left for the next chapter. This section serves as an introductory to the chapter.

The mechanism which describes the way learning and evolutionary search interact is quite different from the previous two models; for example, in the G-P correlation model, learnt behavior is simply replaced by an innate predisposition over generations, whereas in Deacon's model the relationship between learning and evolution is more obscured. To describe this, two important concepts forming a kernel of this mechanism have to be briefly explained. The first is *extragenetic inheritance*[3]. In a nutshell, extragenetic inheritance is a product of organisms' activities that consequently modify surrounding environments which are non-genetically transmitted to the next generation. A range of instances that exhibit non-genetic heredity have been reported at various levels, beginning from the level of cell divisions to human cultural transmissions. However, the most related case of extragenetic inheritance

[3]For some people, this is known as "epigenetic inheritance". However, the term "epigenetic" often denotes a certain short period in a developmental stage of organisms, and subsequently epigenetic inheritance often specifically refers to some chemical inheritance in a cell or so. Thus, here, to avoid an unnecessary confusion, we stick to the term extragenetic inheritance to describe behavioral, social, and cultural evolution.

here is apparently cultural inheritance, since cultural inheritances are creations of higher cognitive abilities. Dawkins (1982) and others have emphasized that most occurrences of cultural behavior do not halt within a generation, but are inherited by later descendants. Although it is truism that such cultural inheritances are ultimately governed by genes (as physical DNA or RNA base sequences are ultimately responsible for the existence of any biotic creatures), their causal relationships are largely indirect and possible genic effects are negligibly small in the comparison with other factors. Cultural inheritance is, thus, considered to have its own channel through which its information is passed to next generation. Dawkins (1976) developed this idea and created the famous "*meme theory*". This higher order extragenetic inheritance plays an important role in Deacon's model. He terms it "*social transmission*".

The second factor of the mechanism is *bilateral interactions* of organisms and their environments. Under the 'standard' concept of the Modern Synthesis, organisms are the ones exposed to continuous threats from their environment. Underlying this concept, it is undoubtedly true that such environments are considered to be static across a number of generations of the organisms. On the other hand, in Deacon's model, some of organisms' activities are considered to act as functions to some parameters in environments; their behavior actively modifies a given environment. This is a crucial conceptual leap in evolutionary theory. If the environment is a subject to be modified by organisms' activities, it is naturally conceivable that as a consequence of the modification, the selective pressure of the organisms may also be deformed to some extent.

Of course, this type of bilateralism has been considered since Darwin expressed his theory of evolution. The best example is indeed found in Darwin's (1881) work. Darwin himself described the way in which earthworms modify compositions of the soil where they live (*i.e.*, the crucial part of their ecology) by eating the soil and excreting the 'digested' soil. For the later generation (partially for the current generation too), the modified soil becomes part of the ecology. However, until recently this type of concept had not been seriously reconciled. Recently, this negligence of bilateral interaction has been gradually reconsidered in biology (*e.g.*, Lewontin 1983). This point will be discussed in the next chapter.

These two systems (*i.e.*, extragenetic inheritance and bilateral interactions) combined with learning formulate the core of Deacon's model. In summary, the basic logic is briefly described as follows: Learning produces an extragenetic inheritance such as cultural inheritance. The inherited trait subsequently formulates a part of the ecology; it modifies certain aspects of the surrounding environment. The

modified environment then produces a new selective pressure. Consequently, this new selective pressure triggers a canalization process which had not existed before.

Following the classification of the three stages of the Baldwin effect introduced by Simpson (1953), let us look at this in a schematic way[4]. In Stage 1, some 'smart' individuals, who can acquire a certain behavior (including cognitive abilities), enjoy a good breeding chance, and subsequently the proportion of such smart individuals in the population increases. Unlike the case of Baldwin's original mechanism, the selective pressure is not necessarily extremely harsh. Rather, like the case of normal adaptation, initially successful learners slowly increase their offspring over the generations. Secondly, Deacon conceives that such inherited learnt behavior itself modifies the ecological condition of the population. In other words, this means that the population not only receives feedbacks from the environment by means of natural selection, but is also equipped with their own feedback mechanism to their environment.

The transition from Stage 2 to Stage 3 takes place because of the new ecological condition. During Stage 2, the environment starts threatening the population in different ways; different types of selective pressure arise as a consequence of an environmental modification. When the environment starts to change, it is assumed that a non-trivial number of individuals in the population are already successful learners who enjoy their learnt behavior, because the environmental modification is a collective result of the learners' activities. The transition is triggered by this new selective pressure; this time, among successful learners, a new competition takes place. The modified environment provides a new type of selection pressure which may or may not be related to the original pressure.

Note that the direct transition from Stage 1 to Stage 3 would hardly occur since the genotype that pulls the population from Stage 2 to Stage 3 is only advantageous in Stage 2 and Stage 3 but not in Stage 1. In other words, it is the collective effect of the behavior produced by the learning individuals in Stage 1 that produces the selective pressure existing in Stage 2. Thus, reaching Stage 2 naturally leads the population to move onto Stage 3.

One might raise the question as to whether this can be indeed called a model of the Baldwin effect. Bearing the Simpsonian model in mind, Deacon's model is certainly unorthodox. This would also be one of the major reasons that Deacon's model has not been reckoned as a model of the Baldwin effect. Basically the doubt revolves around the belittlement of learning and the indefiniteness of the cause-consequence relationship in the model of the Baldwin effect. As one might have

[4]This analysis is first introduced by Godfrey-Smith (2003).

already noticed, the significant part of the model is implementable without invoking learning. This literally means that learning is not a necessary condition for the model to work. The important part of the model is behavior which is able to modify the environment, and subsequently create an extragenetic inheritance (then, which changes the organisms' selective pressure in later generations). The role of learning, if any, is rather indirect. Or put in a different way, learning itself does not have a particular role in the model. In this type of model, learning is rather taken for granted as often behavior engraving an extragenetic inheritance is socially transmitted. One of Deacon's examples about the evolutionary causal relationship between a culture of dairy consumption and lactose tolerance is also described by Durham (1991) without making a particular acknowledgment for the behavior as a learnt one.

This strongly suggests that learning does not directly affect the course of evolution as other models show. This 'basic' type of model has recently garnered wide attention in evolutionary biology. Although some variations of models exist in the field, and they have been independently studied, those models share the basic concept; extragenetic inheritance and bilateral interaction. Susan Oyama and others (*e.g.*, Oyama 1985, Griffiths & Gray 1994) have developed a theory called "*Developmental Systems Theory*" (**DST**). In this theory, they play down the gene-centric point of view –genetic determinism, and put more importance on environmental factors during developmental process. In a similar manner, but stressing bilateralism, Laland and his colleagues have proposed a research program, namely Niche Construction (**NC**, *e.g.*, Odling-Smee *et al.* 2003). While in DST, extragenetic inheritance and bilateral interactions are resigned to secondary roles, NC is a theory which sets the major point in ecological inheritance. A schematic figure is shown from Laland *et al.* (2000, p. 134) in Figure 4.1 (p. 102). Deacon has a similar model to this in his mind when he formulates his own model of the Baldwin effect (*cf.*, Figure 4.2, p. 102 with Figure 4.1). This is the reason that this mode of the Baldwin effect is termed "*Baldwinian Niche Construction*" (**BNC**). A detailed explanation of niche construction itself will be given in Chapter 5.

Deacon's recognition of the Baldwin effect also adds confusion. He never actually uses the term "the Baldwin effect" in the book, instead, he frequently calls "*Baldwinian selection*". Deacon seems to use this term to explain niche construction[5]. Here are two quotations from his explanation of Baldwinian selection and an attached figure (Figure 4.2, p. 102, Taken from Deacon (1997, p. 323)):

[5]In a recent literature (Deacon 2003), he acknowledges that he was not aware of the study of NC, while he was writing *The Symbolic Species* (Deacon 2003).

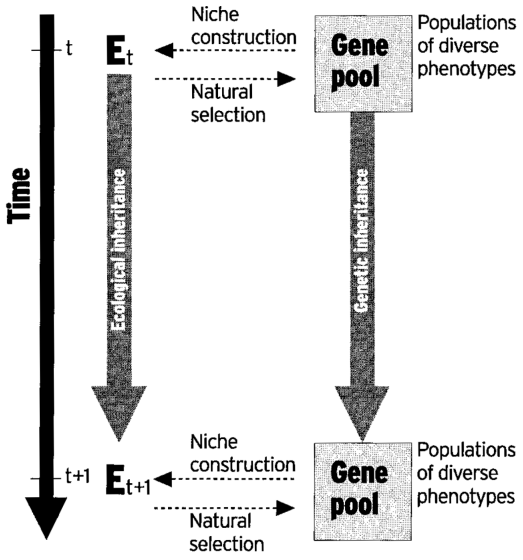

Figure 4.1: Niche Construction

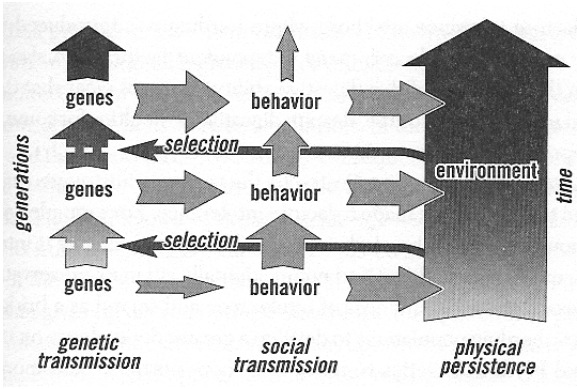

Figure 4.2: Baldwinian Selection

> Baldwin suggested that learning and behavioral flexibility can play a role in amplifying and biasing natural selection because *these abilities enable individuals to modify the context of natural selection that affects their future kin.*
> (Deacon 1997, p. 322: emphasis added)

> Baldwin proposed that by temporarily adjusting behavior or physiological responses during its lifespan in response to novel conditions, *an animal could produce irreversible changes in the adaptive context of future generations.*
> (Deacon 1997, pp. 322-323: emphasis added)

From these, it is apparent that Deacon's Baldwinian selection is fundamentally the same as niche construction. The examples of Baldwinian selection in the book

also support this. The case of lactose tolerance is introduced to describe Baldwinian selection (Deacon 1997). However, there is no description whatsoever of how the term is related to the Baldwin effect in Simpson's sense. Although lactose tolerance is also taken up by Durham (1991) in detail, he never uses the term "the Baldwin effect" in the context either. Apparently, the story of lactose tolerance is a case of niche construction, but hardly a case of the Baldwin effect.

Deacon also discusses Waddington's genetic assimilation (canalization). In the description, he introduces Waddington's genetic assimilation as a 'related process' of Baldwinian evolution. Deacon argues that genetic assimilation makes flexible adaptive responses become progressively more canalized. Canalization is, he discusses: "*a more genetically and developmentally sophisticated description of an important class of Baldwinian processes*" (Deacon 1997, p. 324). However, Deacon properly denies the overly deterministic idea haunting the concept of genetic assimilation; denial of simple G-P correlation. This description of genetic assimilation and canalization strongly suggests that Deacon *does* understand the basic concept of the Baldwin effect. However, in his denial of genetic-determinism in the canalization process, Deacon relates the (non-deterministic) concept of canalization to evolution of genetically distant traits inaugurated by niche construction. Through niche construction, a trait indirectly triggers a causally distant trait's evolution. If the evolution of the trait somehow indirectly affects the initial trait, then the initial trait is said to be canalized.

From this, it becomes vaguely clear why Deacon terms niche construction "Baldwinian selection". What exactly he intends to explain with the term is most likely a subset of niche construction. In usual niche construction cases, a feedback loop is open; the trait as the consequence of a niche construction inaugurated by the initial trait is causally distant. In other words, causes and consequences in normal niche construction are not locally constrained. Consider, for example, the case of malarial resistance and yam cultivation. The trait of malarial resistance is attributed to completely different genes than yam cultivation. Yet these are distantly connected, mediated by niche construction. The case of lactose tolerance somewhat more obscure. The end result still revolves around the initial trait; the dairy consumption culture triggers evolution of adult lactose tolerance. However, even in this case, the two traits are still of different types.

On the other hand, by introducing the concept of canalization, Deacon attempts to tie the feedback loop. When the trait at the end of this loop is of the same type as the beginning, and it positively supports the functional aspect of the initial trait, such an evolutionary process works increasing ineluctability of the trait –the canalization process. If learning initiates this process, and the resulting trait canalizes the

learning, it is the Baldwin effect. Importantly, as Deacon denies overly deterministic replacement of learnt behavior, the increase of ineluctability is not necessarily caused by a single factor. Rather, what he calls "*constellation*" of different causes would enhance it. However, this would be still the case of the Baldwin effect.

Chapter 5

Niche Construction

5.1 Basics of Niche Construction

5.1.1 Darwinian and Mendelian Theories Revised

There are a few ways to describe the concept of niche construction. However, the spirit of the concept has been concisely described in the explanation of Deacon's formulation of the Baldwin effect. Thus, rather than repeating the same idea in different phrases, in the following sections, the concept will be reviewed in a logical, and historical manner with its sibling theories and concepts.

The idea of niche construction itself is not particularly new; ideas which state changing ecology by organisms' own activities have been periodically articulated since Darwin proposed natural selection. However, within the Modern Synthesis regime, one of the earliest and the most significant articulations of the concept is Lewontin. In the mid-1970s, (Lewontin 1974) started discussing that environments are neither static nor uniform for each individual receiving the environmental conditions. Rather, a possible relationship between organisms and an environment is more interactive and flexible.

A similar point has been briefly discussed in the description of the reaction norm approach in Chapter 2. By using graphs, the approach visually reveals the reaction patterns of organisms under a specific environmental condition. It has been realized that even under the same set of environmental conditions, the patterns of reaction are different in each individual. Although a detailed explanation of the reaction norm approach will be provided later, for the time being, it suffices to point out that the environment is not uniform for organisms.

Another important contribution towards this flexible interaction between organisms and environments was provided by Dawkins. In his early renowned book *The Selfish Gene*, Dawkins (1976) developed the idea of 'the replicator and the vehicle'; the replicator interacts with the environment through its container called a vehicle, and it is this replicator that is the unit of selection. The replicators are fundamentally immortal, though the vehicles are discarded at the end of one's life cycle. The

distinction between the replicator and the vehicle is not confined itself to genes and phenotypic traits, but it can be extended to other entities as long as they match with the definition.

In the book *The Extended Phenotype*, Dawkins (1982) pushed the idea further claiming that phenotypes are not necessarily bounded by physically produced genes. One such example is behavioral traits. Behavior is not physical entity; rather it is manifested by brain activities as it is a cognitive product. However, Dawkins does not halt his speculation merely at behavioral activities. He literally "extended" the concept to outside of the organism; he considered that if genes can change the form and structure of an environment outside the body of the host organism and consequently such a modification contributes to gene replications, the modification or the modified environment itself should be also called a phenotype. In other words, the organisms are capable of manipulating their environments through their phenotypes so that some portion of such an environment itself is able to be recognized as the "*extended phenotype*".

One of the examples is found in the rearing behavior of the so-called "leaf-cutting ant" (*Atta*). The tropical ants reproduce numerous amount of offspring at one time. As such, they require plenty of leaves to feed the offspring. However, tropical leaves are typically hard for the ants, and supplying such leaves to the young offspring is not suitable. What the ants have developed over the generations is not a way of feeding such leaves directly to the offspring, but preparing the leaves suitable for the culture of fungi. Such fungi are eventually cultivated by the ants and fed to their offspring. By using the fungi as part of their extended-phenotype, the ants indirectly feed their offspring.

Although Dawkins' viewpoint on the relationship between organisms and environments provides an excellent insight, his view is still strongly gene-centered. For Dawkins, environments are the entities to be surmounted by organisms; extended phenotypes are advantageous when fighting against a given environment in a more global' sense. Put in a larger context, Dawkins' view of the relationship between the organisms and the environments is, by and large, intact; organisms are, through their evolutionary processes, equipped with better and better traits to fight against the existing environment. The term "extended" represents this point well; manipulating a part of the environment. This view consequently leads to broadening of the concept of organism; extended phenotypes can also 'evolve'.

On the other hand, Lewontin firmly believes that the relation should be more equitable. In his seminal paper, *Gene, Organism and Environment* (Lewontin 1983), for example, the tone of Lewontin's critique towards genetic determinism is harsh. He states that Darwinian biology has seen the relationship between the organisms

and the environments described in the context of the subject-and-object distinction; the object is the organism and the subject is the gene and the external environment. In the same vein, it can be distinguished as causes and effects; the genes and the external environments are causes and organisms are effects. These clear distinctions are, Lewontin argues, motived by a strong desire of biology to be aligned with nineteenth century physics. However, recall that Newton's perfectly deterministic model was overturned in the face of the Special Theory of Relativity; yet under the new theory, the Newtonian mechanism is completely kept in a redescription. Lewontin stresses the same should be applicable for evolutionary biology; the strongly deterministic aspect of Darwinian biology can be replaceable without modifying the entire perspective.

Lewontin emphasizes that although the mighty combination of Darwinian natural selection and the Mendelian heredity mechanism is the most basic foundation of evolutionary biology, it should be revised with a more organism-referent perspective. He states that there are a number of contradictions in nature. There are two main points in his claim regarding this. First, the Mendelian concept of heredity is too strongly inclined to atomism; genes are the cause and the organisms are the effects. This type of genocentric view leads us to view developmental processes of organisms in strong determinacy. Indeed, in population genetics, the most successful field in evolutionary biology, the role of phenotypes are neglected and genes are almost equated to the organisms themselves. As early as the mid-1920's, a brief ten years after Mendel's neglected work was discovered, Thomas H. Morgan, the co-founder of genetics (with Bateson), provided a pithy description of the basic concept as follows:

> Between the characters, that furnish the data for the [Mendelian] theory and the postulated genes, to which the characters are referred, lies the whole field of embryonic development. The theory of the gene, as here formulated, states nothing with respect to the way in which the genes are connected with the end-product or character. The absence of information relating to this interval for genetics... but the fact remains that *the sorting out of the characters in successive generations can be explained at present without reference to the way in which the gene affects the developmental process.*
> (Morgan 1926, p. 26: emphasis added).

Needless to say, there are reasons behind this somewhat radical assumption; in genetics, evolution is seen at the population level. At this level, it is understood as a change of gene frequency in a certain genepool. To be modeled in sensitive

statistical models, genetically highly deterministic representations of organisms have been adopted; under this scheme, any factors which obscure the G-P relationship ought to be excluded from the consideration for the sake of theoretical clarity. Naturally, the typical rebuttal against Lewontin's claim from this school is that epigenetic development is merely the secondary factor in evolution and need not be considered in the model. If so, because of its potentially very complex nature, the inclusion of such a secondary factor hazards our understandings of clear 'cause and effect' in evolutionary study.

However, as we will see in Section 5.2, there are cases in which the non-trivial degree of 'impingement' from environmental factors to the G-P relationship is observed. These cases make harder to retain the above assumption even in weaker forms. Alternatively, Lewontin has proposed that the role of gene should be seen as determining the range of reaction norms (*i.e.*, genes determine a range of reactions against environmental conditions). This significantly reduces the role of the gene as it would set a 'framework' for a given phenotypic trait.

The second revision is regarding Darwin's natural selection. In a standard theory of Darwinian evolution, environments including other organisms are the source of selective pressure that holds theoretical importance. Thus if a modified phenotype caused by a genotypic change contributed to a given organism's adaptation, then the degree of pressure is weakened without modifying the structure of the surrounding environment itself. Modification of the environment takes place on its own behalf and is not affected by the organisms within such an environment.

Under this tenet, organisms are thought of as becoming 'adaptive' to given environmental conditions by natural selection –evolution. The environment is thought of autonomous from the organisms; it preexists before the organisms and poses a problem. The organism that obtains the best solution becomes the most prolific. Through the cycles of this process, the species adapts. However, Lewontin refutes this way of looking at evolution by labeling the view as a "*lock and key*" model of evolution; environments pose problems (*i.e.*, the locks) which would be solved by adaptations (*i.e.*, the keys). In this view, thus, the roles of organisms and environments are undoubtedly separated.

However, environment is not an objective, but a highly subjective entity for organisms. For small organisms, such as water fleas, Brownian motion is a crucial factor, while for a dolphin it hardly becomes even a minor factor. On the other hand, buoyancy is not a major factor for the former, while it is for the latter. This and similar kinds of examples strongly suggest that environments do not exist in autonomy. Lewontin warns that the term "adaptation" has a powerful metaphorical

effect (similar to the term "genetic assimilation"), and one is typically inclined to conceive that environments are definable in a vacuum of living creatures.

This abandonment of the strongly adaptive view of evolution, however, leaves a non-trivial paradox behind; organisms are apparently 'fit' to a given environment. It has been said that extant species are all fit to the current environments. How does one describe this obvious fact of nature? Lewontin proposes that all living creatures somehow 'construct' their own environments by interacting with their surrounding environment. He states:

> What is left out of this adaptive description of organism and environment is the fact, clear to all natural historians, that the environments of organisms are made by the organisms themselves as a consequence of their own life activities... Organisms do not adapt to their environments; they construct them out of the bits and pieces of the external world.
> (Lewontin 1983, pp. 63-64 in the reprint)

Thus, organisms carve their ecology and construct their own environment. Under this view of evolution, both organisms and environments are subjectively treated; in other words, both of them are cause and effect at the same time. Elsewhere, Lewontin nicely summarizes this as "*walking on a trampoline*" (Griffiths & Gray 2001); a pithy metaphor describing a fitness landscape can be modified as organisms moving around.

As noted earlier, this way of looking at organisms and environments is certainly not new, like Darwin's study of earthworms. The example described above is also a case in this scheme. The leaf-cutting ants are infamous for their greediness as they cut so many leaves that their surrounding ecology is non-trivially damaged.

This bilateral view of interactions between organism and environment is often referred to as "constructionism" because of its constructive stance towards 'adaptation' to the environment. The following four points of this constructive process are discussed by Lewontin:

1. Organisms determine what is relevant.
2. Organisms alter the external world as it becomes part of their environment.
3. Organisms transduce the physical signals of the external world.
4. Organisms create a statistical pattern of environment different from the pattern in the external world.

Although the fundamental insights of Dawkins and Lewontin divide the two important evolutionists view of evolution rather sharply, it is intriguing to observe the

similarity and difference between Dawkins' extended phenotype and Lewontin's constructionism. On the one hand, both Dawkins and Lewontin propose the extended concept of the organism-environment interaction; for Dawkins, organisms are capable of *manipulating* some part of the environment while for Lewontin organisms are *modifying* the environment via their own activities. For Dawkins, environments are to be surmounted, while for Lewontin, organisms and environments are inseparable, mutually dependent entities. In other words, Dawkins still retains the somewhat environment-referent view of evolution, Lewontin views it from a more organism-referent point.

Yet, both approaches point out one significant implication, namely extragenetic inheritance and bilateral feedback. As Dawkins famously put it, for example, once beavers construct a dam, it typically sustains for, at least, a few generations. For the offspring, some of such environment conditions become an STE; because the mechanism of inheritance does not rely on genetic properties, this inheritance is different from genetic one. In the constructionist approach, this point is rather naturally suggested; in the case of the leaf-cutting ants, what the ants destroy sustains over the generations. Therefore, in the constructive approach, the bilateral feedback system and extragenetic inheritance are two sides of the same coin.

5.1.2 The Elements of Niche Construction

Lewontin's constructivist approach does not designate a specific theory or model which provides a testable prediction or hypothesis. Rather, it is a basic perspective on evolutionary processes. As such, it should be regarded as an umbrella concept; it embraces different styles of theories as its subsets. For example, DST is one such theory. Oyama and others have been working on this theory for nearly twenty years (*e.g.*, Oyama 1985). As noted earlier, DST is more aligned to Lewontin's revision of Mendelian heredity. The major tenets of DST are as follows (from Oyama *et al.* 2001, p. 2):

1. **Joint Determination by multiple causes** Every trait is produced by the interaction of many developmental resources. The gene/environment dichotomy is only one of many ways to divide up these interactants.
2. **Context Sensitivity and Contingency** The significance of any one cause is contingent upon the state of the rest of the system.
3. **Extragenetic Inheritance** An organism inherits a wide range of resources that interact to construct that organism's life cycle.
4. **Development as Construction** Neither traits nor representations of traits are transmitted to offspring. Instead, traits are made–reconstructed–in development.

5. **Distributed Control** No one type of interactant controls development.
6. **Evolution as Construction** Evolution is not a matter of organisms or populations being molded by their environments, but of organism-environment systems changing over time.

As it saliently appears, DST inherits the core spirits of constructionism; it is also apparent that the main avenue the theory approaches is developmental processes that are the stronghold of Mendelian genetic determinism. Subsequently, although it considers extended inheritance and its influence on environmental modifications, they sit in a secondary position in the theory.

Note, however, that what the term "construct" and its derivative words (*e.g.*, constructivism) denote is implicitly used two different ways in DST and others theories. This dual meaning of the term indeed reflects the fact that the concept itself is somewhat confusingly adopted in two different senses; along the line of Lewontin's revision of Darwinian view of natural selection and Mendelian view of heredity. This is well represented in the above enumerations 4 and 6. For example, in DST, such terms are mainly used in the context of developmental processes (*i.e.*, revision of Mendelian heredity). For instance, Gray (1992) nicely depicts developmental processes as construction, not transmission; traits, blueprints, or potentials are not transmitted across generations. In other words, development-centered theories are thought of as organism-internal reference of constructive processes.

On the other hand, as the name denotes, the theory of Niche Construction aims to be more external. That is, NC reconsiders the Darwinian evolutionary process[1]. In this theoretical approach, the term "construction" delineates organisms influence on their surrounding environment; this constructivist approach puts its primacy on organisms' modification of their own environment. The modified environment, in turn, gives a new selective pressure. This evolutionary process proceeds in reciprocal cycles of environmental modification and selection. NC views such organisms' modifications of the environment (*i.e.*, niche construction) and its returns (*i.e.*, natural selection) as 'feedback processes', namely construction of ecology.

NC is set by F. John Odling-Smee, Kevin N. Laland, and Marcus W. Feldman in the late-90's (Odling-Smee 1988, Odling-Smee 1994, Odling-Smee *et al.* 1996, Laland *et al.* 1996, Laland *et al.* 1999, Laland *et al.* 2000, Laland *et al.* 2001*a*, Laland *et al.* 2001*b*, Odling-Smee *et al.* 2003). Laland *et al.* define niche construction as follows:

[1]Needless to say, construction in the developmental sense is also relevant here, but it is assigned a secondary role in the theory.

> Niche construction occurs when an organism modifies the functional relationship between itself and its environment by actively changing one or more of the factors in its environment, either by physically perturbing these factors at its current address, or by relocating to a different address, thereby exposing itself to different factors.
> (Laland *et al.* 2000, p. 165)

The essence of niche construction is summarized as *Co-definition* and *Co-construction* (Gray 1992):

1. **Co-Definition** Any meaningful description of an internal factor must be environmentally referenced and vice versa.
2. **Co-Construction** An organism's environment plays a role in determining the organism while the organism in turn modifies its environment.

As part of constructivism, NC also puts a theoretical importance on extragenetic inheritance. The theory conceives that some niche-construction organisms may modify the selection environments of their offspring. Thus each generation inherits not only genetic information, but also a legacy of modified selection environments. As a whole, this is the idea of ecological inheritance:

> ...any case in which an organism experiences a modified functional relationship between itself and its environment as a consequence of the niche-constructing activities of either its genetic or ecological ancestors.
> (Laland *et al.* 2001*b*, p. 119)

In contrast to other theories in the constructionism approach, NC makes this second kind of inheritance built in the theory and considers it as one of the key factors of the evolutionary process.

Another important departure from Lewontin's own constructivism is found in its explicit commitment of theoretical consideration of higher-order phenotypic traits; namely behavioral traits, especially human behavior. Instead of just considering purely physical/physiological traits, NC attempts to shed light on the more behavioral side of evolutionary processes. In particular, the theory considers the human psychological, social aspect. Together with this theoretical interest of human behavioral traits, this extragenetic inheritance embraces cultural evolution in its theoretical perspective.

Also, its open-endedness of feedback processes makes the theory different. Lewontin implicitly considers that modification of a given environment consequently changes the selective pressure on the same locus which is responsible for the modification itself. On the other hand, NC leaves open the option that such a modification on

the environment affects to change other selective pressures. In other words, NC considers indirectness of feedback from the selection environment. As in Lewontin's revision of Darwinian natural selection concept, one of the important tenets of constructionism is that phenotypic influences to the selection environments are taken into their accounts of evolutionary consequences. However, such studies usually focus on the evolutionary consequences of the loci that are attributed to the expression of the phenotype.

On the other hand, NC attempts to embrace evolutionary consequences on different loci to their scope too. The cases Deacon raises as his example of the Baldwin effect (*i.e.*, lactose tolerance and malarial resistance) are considered to be of this type, rather than the cases of the Baldwin effect. In both cases, the genes attributed to the dairy consumption behavior and the yam cultivation do not get feedbacks from the ecological consequences of the activities; the genes responsible for lactose tolerance and malarial resistance are the targets of such feedbacks. Recently, Deacon has been advancing the idea that such activities 'unveil' new selective pressures in causally distant traits termed the "unmasking" effect (*e.g.*, Deacon 2003). This point will be briefly discussed in Chapter 10.

5.1.3 Niche Construction and Exaptation

One of the most important aspects of niche construction is that it inherently involves the exaptation process. Gould (1991) concisely defines exaptation as "*features that now enhance fitness, but were not built by natural selection for their current role*" (p. 47). For such features to be 'coopted', some environmental change has to take place. In a standard evolutionary process, such a change 'autonomously' takes place; organisms have no control over, say, climate changes, meteor collisions, or bushfires. However, in the mode of niche construction, because of bilateral feedbacks, the organisms could 'pull the trigger' of such a coopting process. Thus, they are the 'agent' of environmental modifications.

The implication of this exaptation aspect of niche construction is non-trivial. If, with some independent reason, a population has abundant neutral phenotypic variance which is rooted in its genetic diversity, through a niche construction process, exaptive selection may take place. This indicates that genetic operations, such as mutations or recombinations, are fundamentally not required. Since adaptive mutations are thought of highly 'fortuitous' events, normally a possible pace of evolutionary process is assumed to be very slow. However, given the non-mutational evolutionary process via exaptation, the pace of evolution in niche construction can be quite rapid. Especially, if niche construction takes in a behavioral/psychological

domain, this pace may well be even faster. This point will be discussed in Section 6.5.

5.1.4 Types of Niche Construction

Regarding the types created by niche behavior, two are broadly conceivable. The first is called "*external* niche construction". External niche constructions occure where a given organism's behavior physically modifies their external environment. It is external because such a physically modified environment also affects other organisms in different species sharing the same ecology. In other words, such a physical environment is objectively evaluable[2].

On the other hand, some niche constructions are called "*internal*" as such modifications are not recognizable from other species[3]. The most common type of internal niche construction is socio-cultural niche construction. Socio-cultural behavior typically produces a certain protocol or norm to which individuals in a given population are encouraged to conform as the environment. However, such an environment is most likely only meaningful within the same species (or even within the same group); the modified (or created) niche is only meaningful species- or group-internally. This is why such a type of niche construction is labeled internal niche construction.

Another dimension on which to classify niche construction is mostly related to the case of internal niche construction. In a social niche construction, if a niche is created through competitions among members, it is called "*competitive* niche construction". It is often the case that in a competitive niche construction, niches are unstable even if some equilibria are observed. In the extreme case, competitions lead a runaway process. A good example is found in the study of the Evolutionary Iterated Prisoners' Dilemma (**EIPD**). In EIPD, typically members genetically inherit some type of strategy to compete against other members. During one's lifetime, members are competing against each other by determining their next behavior (either cooperating or betraying) based on their strategies and previous results (*i.e.*, the history of their competition). The most well-known strategy is called "*Tit-for-Tat*". This very simple strategy has been proven to be the best strategy. Having said that, tit-for-tat is not invincible; it is known that some strategies completely outperform the strategy. The strength of the strategy is relative and context-sensitive. Therefore, while in the majority of the initially-random populations, typically tit-for-tat individuals evolutionarily become prolific, domination of the strategy allows

[2]Although, as in the description of co-definition, for each species, the effect of the environmental modification may be different.

[3]Note however, in the case of symbiosis, different species will be involved even though it is a case of internal niche construction.

some mutants to invade; they are highly vulnerable to cheaters. Thus individuals whose strategy is prone to be betrayal from the beginning can easily earn high fitness by competing with the tit-for-tat individuals. Since the population is nearly saturated with the tit-for-tat individuals or those who are equipped with similar strategies (*i.e.*, cooperative strategies), betrayers can quickly increase their population. This can be thought of as a punctuation of an equilibrium. However, as the number of such betrayers increases, the whole population's adaptability goes down, once again cooperative individuals gradually dominate the population.

In other cases, competitive niche construction creates a radical runaway process. An example is sexual selection. Once a competitive ground is created, endless competition begins. This is because such competition is context-sensitive; one's fitness is determined by other individuals' abilities but not by an externally referable standard. Subsequently, an arms race begins. This continues until some external factor starts to intervene (*e.g.*, the cost of maintaining the competitive ability becomes too high). In this regard, sexual selection is also a type niche construction, since improvement of a sexual trait produces a new context in which any further improvement on the trait has its meaning. Therefore, sexual selection is a type of internal competitive niche construction.

"*Cooperative* niche constructions" are a type of niche construction in which each member's behavior cooperatively creates the environment. In other words, members have to align their behavior with others in order to increase their fitness. Therefore, such a niche can be considered as a sort of norm or protocol itself. Subsequently, cooperative niche constructions often involve socio-cultural behavior. This naturally leads to the fact that the concept has a close relationship to internal niche construction. Of course, there are some cases where interspecific cooperative niche construction takes place, like cooperative symbioses. Having said that, the majority of cooperative niche constructions are thought of as conspecific as is in internal niche construction.

As opposed to competitive niche construction, one interesting point of cooperative niche construction is that an equilibrium is easily created. Once a protocol (*i.e.*, an internal niche) is set, any dropout will lose his adaptability. Such equilibria are thought of as neutrality. This point will be discussed in the last chapter.

Another axis on which to classify the type of niche construction is regarding the agency of inception of niche constructing process. The first type is *counteractive* niche construction (Laland *et al.* 2000). In this type of niche construction, organisms modify their environments in response to autonomous environmental changes so that they re-establish the adaptive match of their phenotypic features with the modified environment. The second is *inceptive* niche construction. Through their

newly innovated activities, for example, organisms may start to modify the environment that, in turn, puts a previous feature-factor relationship into a new state. This type of inceptive niche construction has a close relationship to the creative ability of new behavior (*i.e.*, innovations).

It is thought that while counteractive niche construction is often found in external niche construction, social, internal niche constructions are mostly categorized as inceptive.

5.2 Epistasis and Plasticity

5.2.1 Plasticity in the Reaction Norm Approach

In Chapter 2, we briefly described plasticity from the reaction norm perspective. Canalization can be grasped as a process narrowing such a reaction norm. Recall that in that perspective, plasticity is conceived as a property of the reaction norm of a genotype. Reaction norm is usually visually represented by simple two-dimensional graphs; it expresses the property of a genotype to produce different phenotypes in different environments. More specifically, it captures plasticity as a function which relates an environmental input to a phenotypic output in a defined phenotype space. These two main causal factors –genetic and environmental factors are often labeled **G** and **E**, respectively.

For a given genotype in any environment, the line plotted on a graph will be flat if environmental factors do not affect the phenotype; whatever the environment, the genotype expresses exactly the same phenotype –no plasticity at all. However, there are differences among genotypes regarding their corresponding phenotypes. These genotypic differences are expressed in the graphs as the widths between the slopes. If such lines are plotted for all possible genotypes, this corresponds to the variability of the genotypes. In Figure 5.1 (p. 117), while no environmental factors give effects (therefore, all lines are flat), differences in genotypes give phenotypic variations (the lines occupy different locations in the graph).

If a given genotype is sensitive to the environmental factors, a slope will be observed on the graph. For example, Figure 5.2 (p. 117) shows a case with no genotypic difference, but only environmental factors providing phenotypic variations (all lines are squashed into one, but a slope is detected). The total difference in phenotypic values of the given genotype defines the range of plasticity of the genotype.

However, in reality, such a simple dichotomy rarely occurs in nature; rather, more complicated interactions between genes and environments are observed. The famous experiments conducted by Clausen *et al.* (1948, 1958) are a classic example of this; these experiments took advantage of a plant called "*Achillea millefolium*".

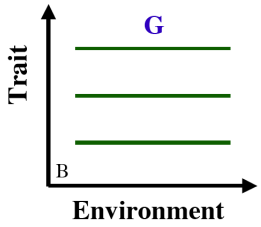

Figure 5.1: Genetic Factors only

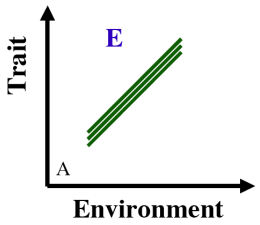

Figure 5.2: Environmental Factors only

Achillea is a plant which can be completely regrown from a piece of its complete form. This means that the regrown plant is genetically identical to the plant from which it is derived. Clausen *et al.* transplanted this plant along a transect in California. Each environment is different in many respects, but the most significant one is altitude. The plants grown in different elevations exhibited interesting results. They found that plants which grew tall at certain elevations were shorter when propagated at other elevations. Moreover, this variation of height at different elevations is not constant across different genotypes. That is, one genotype that grows taller at elevation A would be smaller in elevation B, while another genotype grows smaller at A and taller at B. It is not the case that in one environment, all plants are relatively small, and in the other environment, they grow tall. Figure 5.3 (p. 118) shows a part of the result.

As the experiments of *Achillea* exhibit, the picture of realistic plasticities is more confusing. The results show that something more than **G** and **E** as independent factors exist; some genotype is more sensitive than others to some environment, while in another environment the susceptibility may go in reverse. This proves that individual genotypes idiosyncratically react to a given environmental condition. In the previous graphs, either all lines of genotypes are parallel or squashed. If the slopes are not parallel, it means that genotypes unevenly react to a given environment. This situation is called "*genotype-by-environment interaction*" and is usually abbreviated as $\mathbf{G} \times \mathbf{E}$. The corresponding schematic graph appears in Figure 5.4 (p. 118).

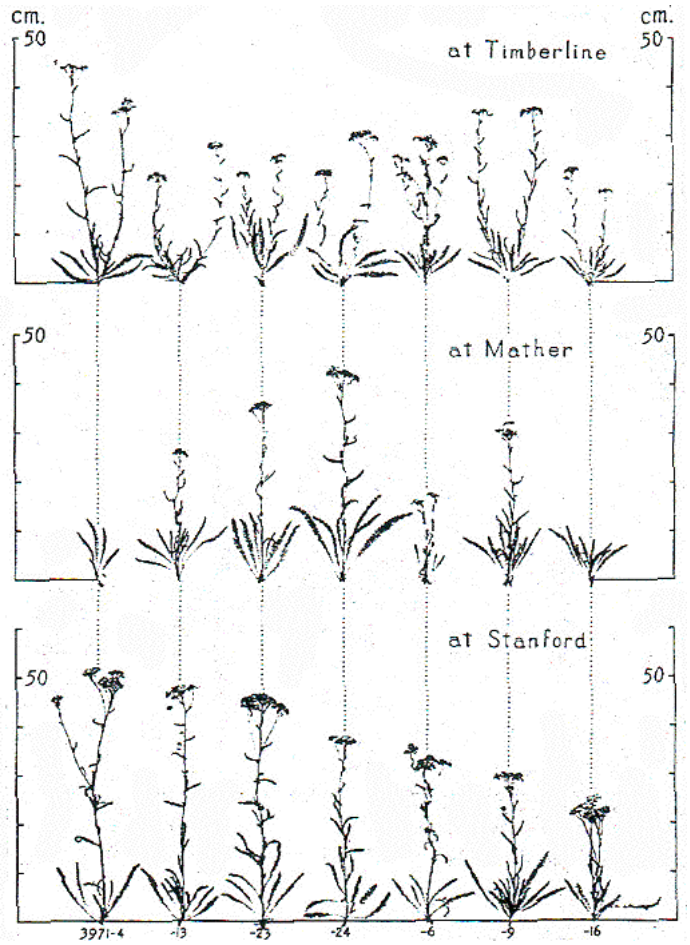

Figure 5.3: Phenotypic responses of *Achillea millefolium* in different elevations

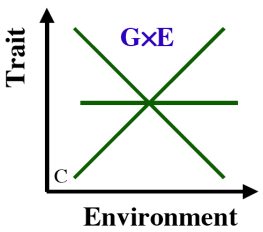

Figure 5.4: Genotype-by-Environment interaction only

It is important to distinguish plasticity at the individual level and at the population level. Consider, for example, Figure 5.4. This shows, at the individual level, different degrees of plasticity. However, a statistical analysis (such as ANOVA Lewontin 1974) will fail to detect genetic differentiation or plasticity in the 'population'. That is, although individual genotypes are different in their plasticity, at the population level, the average trait value of each genotype is the same overall. Thus it is important to differentiate plasticity at the individual and the population level.

Although it is hard to tell from the result of Clausen *et al.*'s experiment, often even under the genotype-by-environment interactions, both **G** and **E** may independently affect reaction norms. In other words, in such a case, while each genotype's reacting pattern to a given environment are different, some general tendencies can be detectable (**E** and/or **G**). When all factors get involved, a possible graph should be as follows. In this case, statistical analyses will detect **G**, **E**, and **G** × **E** (Figure 5.5, p. 119)[4]

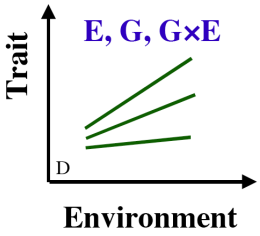

Figure 5.5: All Factors

5.2.2 Epistasis

While a growing body of research on epistasis has recently revealed a complex picture of this non-additive polygeny, here a very simple summary is provided so as to underscore some general properties of epistasis.

One of the most fundamental properties of epistasis is its polygenic aspect. Polygeny is described as a phenomenon in which two or more genes contribute to the expression of a phenotypic trait. Such a phenotypic trait is referred to as a "*polygenic trait*", since a given trait is expressed not only by a single gene, but more than two genes. If such a mechanism is inherited, it is called polygenic inheritance. One of the popular examples of polygeny is cats' coat colors; a cat whose one parent has an all white-coat and the other with an all-black coat does not necessarily have either all-white or all-black. Rather it is often the case that the cat has a number of white spots or black stripes. Thus, the coat colors of cats have a composite nature.

[4]All these three schematic graphs are taken from (Pigliucci n.d.).

Although polygeny is a necessary condition for something to be epistatic, the concept of epistasis is not sufficiently described by it. Mixing up the idea of polygeny with epistasis is a common confusion haunting the concept. The difference between simple polygeny and epistasis is the nature of additivity upon a phenotypic expression. Recall a standard GA model, for instance. In such a model, a genotype, or a set of genes, is used for problem solving. Each genotype is, *as a whole*, a solution to a 'single' problem. In this regard, genotypes in the GA have the nature of polygenic traits; alleles in each locus are typically 'a part of' the solution.

Although genotypes are polygenic, such a model shows a gradual evolutionary process for two important reasons. The first reason is that the objective function of the model sets adaptive differences between different genotypes in a gradual sense. In other words, phenotypic values of genotypes are sufficiently fine-grained; differences between phenotypes are gradual after their fitness values are calculated based on the function. Thus, selection can distinguish differences between the solutions. The second reason is, more importantly, correlations between different levels of genetics. First, differences between genotypes, which are often measured by the Hamming distance, are positively correlated to their phenotypic values (*i.e.* G-P correlation). Secondly, differences of such phenotypic values are positively correlated to differences of their corresponding fitness values. Therefore, in a GA, the two different positive correlations in the different levels are generally assumed. Regarding the first, if two genotypes are close to each other in their Hamming distance, it is also true of their phenotypic values. Besides this, such proximities should be retained between phenotypic values and fitness values. If these relations are retained, a fitness landscape will be a Fujiyama.

This example strikes the point that some polygenic inheritances are considered to be *additive*. On the other hand, the term epistasis is often associated with 'ruggedness' of the fitness landscape. Since apparently in some cases of polygeny, especially additive cases, fitness landscapes have a single, smooth peak, there are some different properties that differentiate simple polygeny and epistasis. In other words, it is conceivable that, though a phenotypic trait has a gradual nature, it does not correspond to the Hamming distances among genotypes. Hence, two very similar genotypes may have quite different phenotypic values and/or fitness values. This is referred to as *non-additive* polygeny. A possible fitness landscape of such a non-additive polygenic inheritance is often rugged. Therefore, the concept of epistasis designates a mechanism of this type of non-additive polygenic system[5].

[5]In some of the literature, epistasis seems to be mistakenly used just for describing rugged fitness landscapes; even non-additive polygeny is not mentioned.

The above description also reveals an important aspect of this non-additive polygeny, namely the levels of epistasis. Suppose that some loci of a genotype are occupied by one of two possible alleles, say a^1 or a^2. Regardless of order, any locus that has a^1 contributes to an increase of a phenotypic value of a given organism retaining the genotype (say, intensity of coat color). Therefore, the phenotypic effects of loci are strictly additive. Consider, then, that the objective function finds the best phenotype at a middle-value phenotype. Genotypes similar to this genotype are linearly assigned good fitness. Thus, the corresponding fitness landscape will be like a normal distribution curve. Although at first glance, this looks like a simple evolutionary mechanism, it is indeed an instance of epistasis; one locus' contribution to fitness depends on how many other loci have a^1 (or a^2) alleles in a given genotype. If the phenotypic value is below the optimum, a^1 in a given locus positively contributes to the fitness, while above optimum, this works negatively. This is called "*epistasis for fitness*". As Brodie Ⅲ (2000) states, if selection is nonlinear, non-additive fitness effects appear in the loci while they exhibit additive phenotypic effects.

The other level is called "*epistasis for phenotype*". A phenotypic expression or value is non-additively affected by two or more genes. Normally, this is often understood as what epistasis means. The coat colors of cats express epistasis at this level. This is the level where G-P correlation is deteriorated.

Another property of epistasis concerns the difference between the individual and the population. As plasticity in an individual and population is different, epistasis is significantly different at the individual level and the population level. At the individual level, it is called "physiological" (or "mechanistic" or "physical") epistasis, and at the population level, "statistical" (or "populational") epistasis. As in the case of epistasis for phenotype, physiological epistasis is our understanding of normal epistasis; two or more genes in different loci non-additively determine a phenotypic value. On the other hand, the concept of statistical epistasis depends on allelic frequencies; at some allelic frequencies, epistasis strongly appears, while in different frequencies, it may be nearly absent even if exactly the same genes are involved. This is precisely because epistatic effect is determined by interactions between multiple alleles in different loci. Consider an epistatic effect produced by two loci. If one of the loci is set as a focal locus, the other locus can be considered as "genetic background". A specific allele in the background locus determines the effect of the focal locus; different alleles have different effects even with the same allele on the focal locus. However, if the frequency of a specific allele on the background locus is extreme (say, the same allele almost always appears on the locus), it is almost the same thing as the phenotypic trait being a linear function of the focal

locus. As allelic frequency is only measurable at the population level, this form of epistasis is only conceivable populationally. Thus while physiological epistasis specifies the 'range' of epistatic effect, statistical epistasis designates 'transferability' of physiological epistasis onto the population level.

Finally, if epistasis exists among genes, it is natural that "pleiotropy" also exists. Pleiotropy, in a very crude form, means that one gene contributes to express more than one phenotypic character. Thus, one gene in a model will affect an expression of one phenotypic trait, but also will determine other traits. One example occurs if cats are odd-eyed with one blue and one yellow eye. All odd-eyed cats have an all-white coat and often are deaf only in the ear on the same side of the head as the blue eye. The exact cause of this is still unclear, but it has recently been proposed that the cause for both lack of pigment and deafness lies with the gene involved during early embryogenesis in controlling the development.

5.2.3 Epistasis and Norm of Reaction

Given the above descriptions of both the reaction norm approach and epistasis, in the following few sections, more complex genetic interactions are considered.

While $\mathbf{G}$, $\mathbf{E}$, and $\mathbf{G} \times \mathbf{E}$ are the standard notation in a reaction norm, recently, along the lines of the growing attention towards epistatic effects on evolutionary processes, yet another concept has been proposed.

Consider epistasis in the reaction norm approach. Recall that in that approach, phenotypic values are plotted against different environmental factors. If such genotypes are plastic, there should be some detectable slopes in a graph. If genotypes unevenly react to a given environment, it means that there are genotype-by-environment interactions.

Epistasis can be also analogously considered within this scheme. In statistic epistasis, genetic backgrounds determine how much of a physiological epistatic effect can actually appear in nature. In other words, when all possible backgrounds are considered, possible norms of reaction (in other words, physiological epistasis) are identified. If these genetic backgrounds are plotted as if environmental conditions were plotted on a norm of reaction graph, then the range of physiological epistatic effects is interpretable as "plasticity" against the genetic backgrounds.

Similar to a standard norm of reaction graph, additive or non-additive nature of epistasis is recognizable from patterns of reaction. However, the way of detecting non-additive, epistatic effect is slightly different from standard norms of reactions; if a given locus has a simple additive nature to its 'partner' loci, all lines in a graph should be parallel to each other. On the other hand, if the locus is non-additive, the lines are not parallel due to different polygenic reactions with other loci.

In effect, application of the reaction norm approach sheds light on the context-dependent aspect of epistasis. By putting the rest of the loci in the background, the reaction norm view of epistasis emphasizes a possible range of reaction pattern of the given locus. Although epistatic effect is normally considered only within the genotype of the same organism, there is no obvious reason to release this condition so that genotypes in other organisms can be taken into consideration; the introduction of the idea of "genetic background" in epistatic reaction certainly helps the development of this extension of epistasis. In the next section, this concept is examined.

5.2.4 Indirect Genetic Effects and $\mathbf{G} \times \mathbf{G}$

Usually, in the reaction norm approach, environmental conditions taken into account are exclusively abiotic. The above section shows that polygeny and epistasis, however, can also be considered within the same scheme. As these are genetic (*i.e.*, biotic) properties, it is clear that the reaction norm approach can be fundamentally flexible to apply to this type of biotic conditions. Since polygenic or epistatic properties are genetic, the reaction norm approach in this field (*i.e.*, polygeny and epistasis) deals with effects of genetic interactions on development.

A little further investigation, however, provides an interesting possibility of augmenting the power of this approach. So far, our concern with genetic background, instead of environmental factors, is confined to a single organism; polygenic inheritance, including epistasis, only concerns intragenomic cases. However, by focusing on the idea of 'genetic background', combined with the reaction norm view in polygenic properties, it is easy to conceive that such a background is also externally formed. In other words, it enables us to consider other organisms' effects on the focal individual's developmental processes, instead of intragenomic interactions such as polygenic inheritances. This type of interacting individuals is grasped as genetic influences on other genes in another individual; genes of an individual phenotypically affect another individual. As other organisms' influences do not directly influence the focal individual, this type of interaction is called "*Indirect Genetic Effects*" (**IGEs**, Wolf 2000); the genetic effects of the trait are produced in different organisms but not in the individual whose phenotype is measured. More precisely, it is indirect, since the focal individual's genotype is affected by the environment produced by the genotype of another individual. This contrasts with organism-internal gene effects that act on the phenotypes of the focal organism (*i.e.*, "Direct Genetic Effects"). Interactions that create such effects are called "*genotype-by-genotype*" interactions ($\mathbf{G} \times \mathbf{G}$).

One of the basic considerations of IGEs which dissects the concept into two subcategories is regarding the relationship between IGEs and phenotypic traits that create such effects. That is, how specific IGEs are created from phenotypic traits; organisms' traits individually serve the focal organism's environment. The first type is when individual traits themselves create IGEs independently. Therefore, in this type of IGE, there is a relatively straightforward relationship between a given trait and influences on the focal individual. Consequently, the contribution of such a trait in such IGEs is comparatively easy to measure.

The second type is when different phenotypic traits collectively serve as a single indirect effect. This type of IGE mediation is called "*performance*". Performance is thought of as a collective trait which influences the focal organism's environment. While the contribution of IGEs created by individual traits is directly measurable, the contribution of a performance might be hardly discernible in each individual trait of which the performance itself is composed.

However, this dissection is not necessarily exclusive; in some cases, by focusing on a particular trait which forms a performance with other traits, it is possible to perform an experiment that partitions individual effects of individual traits. For example, maternal care in a broad sense is thought of as a case of performance; various styles of parental treatment may exist. But as a whole, they affect an infant's development. However, it is also true that one can consider *a specific style* of maternal care. For instance, in mammals, mothers provide milk for infants. The amount of milk production significantly affects the early developmental process of infants. Although such IGEs can be considered in the context of the broader sense, it is certainly natural to study the trait as an independent IGE. Therefore, the distinction of individual and performance IGEs has a somewhat theory-laden aspect. Having said that, this distinction between individual traits and performance is useful especially when one considers social contexts where cultural interactions are mostly attributed to multiple traits.

The distinction also sheds light on the epistatic aspect of IGEs. First of all, all IGEs are fundamentally polygenic as different genes are involved to express a given phenotypic trait. The question which follows is whether or not such IGEs have a non-additive, epistatic nature. It is true that most current case studies of IGEs concern simple additive instances (Wolf *et al.* 1998). However, this is mainly because of practical reasons. Rather, it is known that genotypic interactions often exhibit non-additive effects. This is particularly true when IGEs in performance are considered. In such cases, a given genotype's expression of a phenotype depends on non-additive interactions with other genes; it is thought of as a type of epistasis, even though such interacting genes are outside of the focal individual. To contrast

this with $\mathbf{G} \times \mathbf{G}$ epistasis, the conventional type of epistasis is called "*intragenomic epistasis*" (Wolf 2000).

By and large, IGEs have two different modes of contribution to evolutionary processes. The first type is to influence the G-P mapping relationship. Although this is somewhat obvious, it is clearly an important type of contribution; as IGEs are an extra source of environmental conditions, they consequently change a possible G-P mapping relationship. In a standard model, the expression of phenotype is determined by genetic factors with abiotic environmental factors. Under such a condition, the possible contribution of genes in a phenotypic expression is, in principle, statistically measurable; in a nutshell, it is measured as the covariance between phenotype and genotype. On the other hand, when IGEs are involved, the contribution of genes to the phenotypic expression is effectively weakened in a relative sense. That is, being an extra environmental factor, especially as an STE, genotypes in other individuals influence the focal individual's phenotypic expression. Therefore, in contrast to the standard condition, epigenetic development becomes, to some extent, more contingent as IGEs are neither static nor internally manipulatable factors.

This mode of contribution has slightly different impacts depending on the types of IGEs. When an IGE takes place within a family lineage, it means that IGEs are parallel to genetic inheritance; what one inherits comes from the same source as what indirectly affects you. Thus there is a covariance between what you inherit and what you experience. If IGEs work positively, such effects enhance the differences of phenotypic values compared to the case where no IGEs exist. This effectively magnifies genetic differences between different genotypes which originally did not exist; as selection works on such differences as usual, it subsequently accelerates the pace of evolutionary processes.

A typical example of this is maternal effect[6]. Parental care is especially common among mammals. By providing such care, young infants are able to survive or grow healthily in otherwise harsh environments. Thus, as parental care alters environments that the infants experience, they are considered to be one type of IGEs. Suppose that a maternal trait positively affects the same trait in the offspring. In this case, the covariance between the genotypic value and the phenotypic value increases. Therefore, in contrast to non-IGE parent-offspring relationships, such IGEs modify the otherwise straightforward relationship between a genotype and a phenotype.

[6]It is a relatively well studied subject of IGEs (Wolf *et al.* 1998)

This modification of G-P relationship by IGEs is also conceivable in more general cases. When what IGEs transmit and the standard inheritance hands on are genetically different, it exhibits a slightly different influencing process on the relationship. Unrelated individuals' traits influence the focal individual. In this case, since there is no direct relationship between genes inherited and the environment the focal individual experiences, the relationship between genes inherited and the environment experienced does not exist. Instead, by experiencing the social environment which is created by an aggregation of such individuals, genotype and phenotype form a positive feedback. For example, if one's level of aggression influences other individuals' aggression, and vice versa, the relationship between the genetic value that is responsible for the aggression and the phenotypic value that is the level of aggression is quickly accentuated (Wolf *et al.* 1998). Therefore, like a single lineage case, this type of 'social' IGE also works to enhance genetic differences at the face of natural selection.

The second mode of contribution of IGEs to evolutionary processes is an even more indirect one. Because IGEs on a given focal individual work as one of the primal environmental components of the individual, different IGEs lead to different phenotypic responses. In other words, alterations of indirect genetic contributions also modify phenotypic values that constitute yet another IGE in the next generation. If such IGEs are cross-generational, and selection takes place on such phenotypic values, a further alteration to the indirect genetic contribution can possibly occur. As a result, this genetically-based environment itself is sensitive to the previous generation's genotypes. Therefore, IGEs themselves effectively make environments evolve. While the concept of the evolving environment effect sounds close to the modification of a G-P relationship, especially the case of unrelated individuals, it can certainly be considered as an independent mode of contribution.

The notion of modification on a G-P relationship describes the fact that the degree of linkage between a genotype and a phenotype is modified by inserting IGEs. On the other hand, the concept of the evolving environment captures the point that IGEs are susceptible to their own creations, as they serve as an environmental factor. The example of aggression is also useful here. Suppose that the level of one's aggression is somewhat controlled by a ritualized behavior (*e.g.*, some display pattern). An evolutionary change of this ritualized behavior triggers a change in the level of aggression in a population. If selection on such behavior is systematically related to one's physical aggression controlled somewhat by the behavior itself, then

the evolving environment effect forms strong positive feedback in the evolutionary process[7].

The evolving environment can be created even when different species produce IGEs. Symbiotic coevolution belongs to this type. While this type of inter-species' IGE is interesting from a perspective of coevolution, as far as language evolution is concerned, the most important type of IGE is still conspecific interactions. Indeed, even putting aside the consideration of the linguistic aspect from our immediate interest, IGEs and the evolving environment in conspecifics still provide important foundations of cultural evolution. First of all, IGEs among unrelated individuals create 'social' environments; such environments are created by interactions of phenotypes and influence a formation of individual phenotypes. Secondly, as cross-generational IGEs create the evolving environment, a portion of environmental factors are inherited. As a result, traits which do not even have a genetic basis can evolve across generations.

Another important aspect of the evolving environment is that such cross-generational changes in an environment often produce evolutionary 'time-lags' or momenta; because a part of the environment that individuals experience is created by the previous generation's genotypes (or phenotypes), a selective pressure in a given generation does not necessarily reflect a precise problem that the abiotic environment poses. This is because IGEs modify the current environment so that other factors of the environment are weakened with regard to their impacts on selection in a relative sense. Therefore, for example, with some reason, abiotic factors changing in a given time *t*, may not take effect immediately because IGEs from the previous generation overshadow the factors, *t-1*. A series of idealized simulations are shown in Figure 5.6 (p. 128; taken from Wolf *et al.* (1998)). In the figure, evolutionary trajectories show the mean value of a maternal trait that affects the expression of the same trait in the offspring. Directional selection was applied for the first eight generations. The generation where selection ceased is indicated by an arrow. (a) Shows the case for no maternal effect, (b) shows the case for a positive maternal effect, and (c) shows the case for a negative maternal effect. (a) shows when no IGEs take place, evolution immediately stops. However, when IGEs are involved (b, and c), even under no selective pressure, the evolutionary process still continues[8].

This type of momentum also produces somewhat counterintuitive evolutionary responses. As in the last graph, if the evolving environment has negative correlation with direct genetic inheritance, it may consequently produce temporal maladaptive

[7]Note that in the G-P relationship, unrelated individual interactions lead to a positive feedback in a single generation, but not necessarily across generations.

[8]For a detailed description, refer Wolf *et al.* (1998)

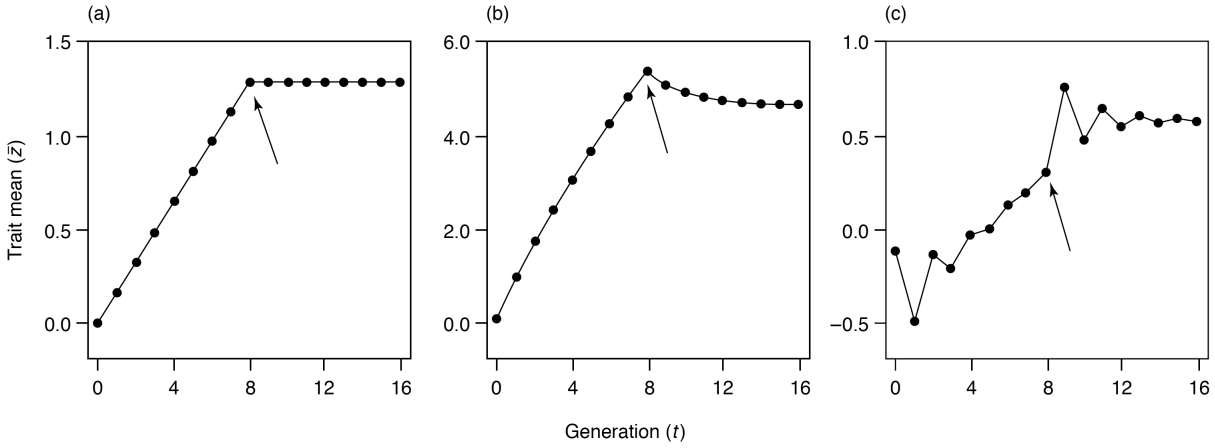

Figure 5.6: "Time-lag" effects

evolution. This is clearly shown in the last graph; although the general direction of evolution is driven by the directional selective pressure, brief maladaptations are observable. As a consequence, an oscillation pattern is formed. This clearly shows that such maladaptations are temporal as such a trend eventually diminishes due to the fact that even such IGEs are a subject of overall evolutionary processes.

5.3 IGEs, $\mathbf{G} \times \mathbf{G}$, and Niche Construction

In the last few sections, some concepts associating with plasticity have been discussed. The reaction norm approach gives the visual interpretation of plasticity. Epistatic effects of genes in phenotypic expressions are also conceivable within this scheme; set a focal gene, other genes which affect the same phenotypic expression are thought of the 'background' of the focal gene. Although the idea primarily targets intragenomic processes (*i.e.*, the intragenomic epistasis), this concept is also easily applicable to considering more complicated processes; when two or more individuals interact and consequently, one affects the other's epigenetic development, it can be said that the former genes indirectly affect the latter genes. Since these indirect interactions of genes exhibit a primal similarity to standard polygenic processes (including epistasis), they can be also considered within the reaction norm approach. Therefore, from this it appears that the reaction norm approach is not only a tool for visualizing plasticity, but also a tool for considering plasticity at different levels. Across the different levels of plasticity, it consistently provides a pivotal view for plasticity; by focusing on a focal gene (or set of genes) and putting other genes as background environmental factors, it enables one to view plasticities from an 'agents and effects' perspective.

Among other levels of polygeny, it is obviously IGEs under the reaction norm approach that are most closely related to niche construction. Indeed, IGEs are another way to look at niche construction. In contrast to the theory of niche construction, under the view of IGEs, plastic reactions are more focused. Since environmental factors which affect the reaction norm come from niche constructing processes, IGEs are in a way to view niche construction from individual bases. It is this focusing effect that IGEs under the reaction norm approach show the most power.

First, as a fundamental part of the concept, social relationships are naturally considered in IGEs. IGEs are produced by interactions occurring not within an individual, but a group of individuals. Although the study of IGEs mostly considers a pair of individuals, this is only because such simple cases simply facilitate research. However, IGEs among unrelated individuals often create a social context from which individuals are influenced. This exactly matches with one of the most basic aspects of niche construction, namely co-construction.

Such a social context produced by IGEs also imposes on its members to behave in a certain manner to maintain (or increase) their adaptivity. Therefore, this imposition of conformity is ultimately created by the individual's (or from previous generation's) own behavior. Putting this in a simple form, behavior that creates a social context receives some selective pressure from the norm created by the context. This is apparently equivalent to co-definition in niche construction.

From these, it is safe to state that IGEs by $\mathbf{G} \times \mathbf{G}$ interactions are fundamentally niche construction. However, an obvious difference exists between IGEs and standard niche construction; while a standard notion of niche construction does not normally consider plasticity, it is an essential component of IGEs. This, however, does not mean that niche construction does not consider plasticity *at all*. Rather, by taking it for granted, niche construction puts its utmost effort on investigating the complex interactions between environments and organisms that create a spiral of feedback processes. Therefore, even if plasticity is a vital component of certain behavior that creates a niche, most researchers have not paid great attention to a possible outcome of it. On the other hand, as in the scheme of reaction norm, the study of IGEs provides a larger picture, while such a view might be narrower in a sense that it usually does not concern 'environments' as it pays more attention to the interactions creating such environments.

However, it is not the only difference between IGEs and niche construction. An even more crucial difference lies between these two concepts relating to the role of plasticity. It is apparent that the study of IGEs puts more stress on plasticity than does niche construction, as the concept of IGEs have been developed from

the reaction norm approach. Previously, it was shown that the process of the Baldwin (especially canalizing) effect can be described as a narrowing-down process of the range of a reaction norm (*see* Ancel 1999). There is no reason that this cannot be applied to IGEs; they also deal with the Baldwin effect. This is a very important insight regarding the relationship between the Baldwin effect and niche construction. It suggests that when IGEs take place, under some conditions the Baldwin effect may occur. In other words, indirect genetic effects are thought of as conceptually embracing both niche construction and the Baldwin effect. IGEs are not exactly equivalent to the concept of niche construction; indeed they more likely cover a 'subset' of niche construction where learning and plasticity are involved. This is the definition of "*Baldwinian Niche Construction*" (**BNC**).

5.4 Baldwin's Social Inheritance

For the majority of researchers in evolutionary study and computer science, the summit of James M. Baldwin's works would be concentrated in *a New Factor in Evolution* (Baldwin 1896*d*). In this work, Baldwin summarized his previous works regarding Organic Selection and extended the idea. As we have seen already, Simpson translated this dated concept to fit the context of the Modern Synthesis, and labeled it the Baldwin effect. Although, for a century, the description of innate predisposition of acquired behavior has certainly been the most influential part of the literature, rather enthusiastic acclaims of the Baldwin effect have overshadowed what Baldwin truly intended to convey. Indeed, in the article, consisting of six sections, the part that describes his own concept of the Baldwin effect comprises just the first two sections. In those sections, Baldwin discussed how Organic Selection would act with phylogenetic evolution. As we have seen, the mechanism employed in the theory is obsolete, even though the overall argument is inspiring enough.

However, it is also true that the overwhelming popularity of the Baldwin effect unfortunately relegates the rest of the work to the background and makes the Baldwin effect the predominant accomplishment. As shown below, in the rest of the article, Baldwin provided a truly progressive concept, the impact of which could be even comparable to the Baldwin effect. To illustrate this, however, we need to remove our stereotypic image of Baldwin and recapture a somewhat more appropriate context of his work. Then it will become clear that his description of the Baldwin effect should be considered not on its own, but it had better be understood alongside his other theoretical beliefs

For example, it has been often neglected among researchers that Baldwin was one of the earliest American experimental psychologists. As such, evolutionary studies are not the primary domain of his research program. Rather, as a part

of his interest in the philosophical aspect of psychology, Baldwin presented several evolutionary studies of psychological development[9]. In one way, therefore, compared to other evolutionists in that era, it could be more natural for him to conceive a possible role of learning in evolution; after all, for Baldwin, psychological abilities are another type of adaptation which could work in the somewhat comparable domain of phylogenetic adaptation.

Baldwin was also the first kind of social psychologist who considered psychological development at the social level. In that era (*i.e.*, the beginning of the last century), evolutionary progressivism had both overtly and covertly influenced social studies; as we can seen in the history of eugenics, various scholars adopted pseudo-Darwinism, and expressed the idea that human societies could be climbing to a moral height by a law of natural selection. This is often called "*social Darwinism*". Baldwin's later works suggest that this type of naturalistic fallacy was haunting him too; he was one of the type of scholars who (mildly) embraced the concept. Indeed, around the year of 1896 (*i.e.*, around the time when *a New Factor in Evolution* was published), Baldwin started publishing his social studies. More precisely, from that period, Baldwin started committing himself to studying psychology in a social context. In this sense, *A New Factor in Evolution* was one of the earliest publications of the sort[10].

Based upon these facts about Baldwin, the omitted aspect of the literature can be reconsidered here. In the later part of the literature, his concept of organic selection was developed under some weak influence from social-Darwinism. While in the earlier part (*i.e.*, where he described organic selection), Baldwin discussed learning in an individual sense, later he speculated on 'social' learning. He considered that 'ontogenetic modifications' (*i.e.*, learning) are used to imitate other members in a given environment. Baldwin called this type of imitation "*gregarious*" in animals and "*social*" in human beings. Baldwin reflected on these 'social learning' processes according to his theory of organic selection; he considered that 'organic' adaptation would be enhanced by this social learning. Baldwin wrote:

> In all the higher reaches of development, we find certain co-operative or "social" processes which directly supplement or add to the individual's

[9]This indicates, therefore, that Baldwin is also one of the earliest scholars in the field of evolutionary psychology.

[10]Interestingly, this indeed seems to be the turning point; from this period, Baldwin seemed to give up the idea of Organic Selection. This is apparent from his work after *a New Factor in Evolution*. Moreover, Baldwin did not even develop the concept much further after the article was published. This is a somewhat perplexing fact as the longevity of his name owes much to the concept.

> private adaptations.
> (Baldwin 1896*d*, p. 66 in the reprint)

This social concept in learning is an important development of his theory; social learning is a non-trivial conceptual leap from individual learning. It is not merely a difference of, say, the number of learners or the size of such processes. Learning something *socially* involves a fundamentally different mode of learning from "individual learning", especially in the case of imitative learning (*cf.* Cangelosi & Harnad 2002). Baldwin termed this mode of learning and its process "*Social Heredity*" (**SH**). This concept shows a surprising similarity to the modern theory of niche construction. Here, with some quotations from the literature, his concept of SH is examined.

First, Baldwin conceived that social learning typically exhibits population internal references:

> [I]t is evident that other living creatures constitute part of the environment of each, and many neuro-genetic and psycho-genetic accommodations have reference to or involve these other creatures.
> (Baldwin 1896*d*, pp. 66-67 in the reprint)

From an evolutionary perspective, this organism-referent aspect of social learning gives an important insight. Namely, if social learning is adaptive and depends on organism-referent cooperation and imitations (as Baldwin argued), the process itself would subsequently shape the selective environment. In other words, the social learning is co-constructing the organisms' environment, and co-defining evolutionary selection. Recall that co-definition & co-construction are the two factors of bilateral-feedback systems. The following brief quotation captures this point:

> [Social Heredity tends] *to produce adaptations which depend upon social coöperation; thus variations in the direction of sociality are selected and made determinate.*
> (Baldwin 1896*d*, p. 67 in the reprint: original emphasis)

Baldwin conceived the organism-referent concept not only in an ontogenetic timescale, but also in the context of phylogeny. Indeed, he stressed that social learning is a hereditary process. In other words, social learning is an extragenetic process as it supports organisms' adaptations, keeps such organisms alive, and subsequently secures their lineage.

> It is a means of extra-genetic transmission from generation to generation. It is really a form of heredity because (1) *it is a handing down of physical functions*, while it is not physical heredity. It is entitled to

> be called heredity for the further reason (2) that *it directly influences physical heredity in the way mentioned*, *i.e.*, it keeps alive variations, thus sets the direction of ontogenetic adaptation, thereby influences the direction of the available congenital variations of the next generation, and so determines phylogenetic development.
> (Baldwin 1896*d*, p. 67 in the reprint: original emphasis)

From these quotations, it is apparent that Baldwin's SH indeed embraces the two important factors of niche construction; *bilateral feedbacks* and *extended inheritance*. Moreover, as is found in the later part of the quotation above (*i.e.*, in part (2)), Baldwin discussed the idea that this process of SH affects phylogenetic evolution; SH shapes the direction of a phylogenetic evolutionary process. In summary, in SH, Baldwin expressed at least the following factors as essential to niche construction; bilateral feedbacks (co-construction & co-definition), extragenetic inheritance, and possible impacts of these factors on phylogenetic evolution. These clearly reveal that SH shows surprising similarities to the modern concept of niche construction.

Note that, even more surprisingly, SH encapsulates more advanced conceptions regarding the relationship between learning and niche construction; the concept of SH stands entirely upon the theory of Organic Selection. In other words, Baldwin conceived SH as a part of the theory of Organic Selection. Indeed, this point is concisely summarized in the literature; "*It is a form of Organic Selection but it deserves a special name because of its special way of operation*" (Baldwin 1896*d*, p. 78 in the reprint). Therefore, for him, niche construction would not be conceivable without some form of acquisition, which appears in the description of Organic Selection. As a result of this 'learning-oriented' formulation of the theory, SH is not only arguing for a simple case of niche construction, but also dealing with a further effect of the process. Regarding bilateral feedbacks, for instance, Baldwin argued that social learning involves other members in a population or in the same habitat as a referent of imitations. This may subsequently create selective processes within the population (*i.e.*, intrapopulation competitions) where previously none existed. According to Baldwin, this is because individuals have differences in their abilities of social learning. Consequently, previously hidden qualitative differences among individuals are exposed to natural selection.

> It is really heredity, since it influences the direction of phylogenetic variation by keeping socially adaptive creatures alive while others which do not adapt themselves in this way are cut off.
> (Baldwin 1896*d*, p. 78 in the reprint)

Subsequently SH is more comparable to Deacon's basic formulation of the Baldwin effect. As we have seen, while Deacon somewhat mysteriously incorporates the Baldwin effect into niche construction, the standard theory of niche construction does not explicitly require plasticity or learning as an essential ingredient for its recipe. Then, this means that, not only is Baldwin the first person who conceived "the Baldwin effect", but also he himself is actually the pioneer of BNC.

Moreover, there are two more important differences between modern niche construction and SH. First, SH is a concept of cooperative behavior and/or learning. Often, the theory of NC or similar theories consider scenarios of competitive niche constructions; similar to sexual selection. Classically, this type of competitive scenario produces various analyses such as the 'Red Queen Hypothesis' and 'Lotka-Volterra systems'. In such cases, evolutionary processes tend to produce extreme results. On the other hand, SH only considers acquisition of socially cooperative behavior (*i.e.*, cooperative niche constructions).

Secondly, SH is based on an assumption of strongly organism-referent, population-internal processes. That is, while niche construction considers adaptations involving co-construction between a given population and other organisms or abiotic factors (*i.e.*, external niche constructions), SH mostly depends on population-internal adaptations (*i.e.*, internal niche constructions). Baldwin considered that imitation is a crucial form of social learning. Of course, a possible adaptive value of learnt behavior through social learning itself should be determined by a somewhat external measure (*e.g.*, how such behavior is useful when faced with a given danger). Yet, success of such learning mostly depends on population-internal factors.

Finally, Baldwin commented that because of the importance of social transmission in human societies, as opposed to other organisms, human beings might keep the plasticity of social learning. He stated:

> In the animals, the social transmission seems to be mainly useful as enabling a species to get instincts slowly in determinate directions, by keeping off the operation of natural selection. Social Heredity is then the lessor factor; it serves Biological Heredity. But in man, the reverse. Social transmission is the important factor, and the congenital equipment of instincts is actually broken up in order to allow the plasticity which the human being's social learning requires him to have.
> (Baldwin 1896*d*, p. 69 in the reprint)

This comment is indeed interesting; as we have seen, within the context of the modern Baldwin effect, computer scientists (*e.g.*, Mayley 1996*a*) have discovered that the degree of canalization is determined by the cost of learning. Although

Baldwin himself did not comment on this, the above quotation proves that there was already a similar, but more niche-construction-oriented idea about the balancing process between ineluctability and learning in his mind.

In summary, Baldwin's SH exhibits striking similarities to niche construction. Moreover, as a part of his theory of Organic Selection, Baldwin envisaged this hypothesis under the 'Baldwinian' hypothesis. Subsequently, Baldwin's own theory of niche construction is, in one way, more advanced and closer to the concept of BNC, rather than the normal type of niche construction being developed. Here, his own excellent summary of Social Heredity in the context of phylogenetic evolution is provided:

> [T]here is *natural heredity* by which variations are congenitally transmitted with original endowment, and there is '*social heredity*' by which functions socially acquired... are also socially transmitted. The one is phylogenetic; the other ontogenetic. But these two lines of hereditary influence are not separate nor uninfluential on each other. Congenital variations, on the one hand, are kept alive and made effective by their conscious use for intelligent and imitative adaptations in the life of the individual; and, on the other hand, intelligent and imitative adaptations become congenital by further progress and refinement of variation in the same lines of functions as those which their acquisition by the individual called into play.
> (Baldwin 1896*d*, pp. 69-70 in the reprint: original emphasis)

However we should be cautious about the fact that the idea of SH was a product of his interpretation of social Darwinism. Subsequently, there might be fundamental differences between Baldwinian niche construction and Baldwin's original Social Heredity. Moreover, as in Organic Selection, it is undeniable that Baldwin's theories are, from the modern biological perspective, obsolete. Even considering these negative factors, however, SH is still very appealing. After all, his conception is inspiring enough to be noted and worth committing further investigations.

Chapter 6

Baldwinian Niche Construction

As apparent from its name, Baldwinian niche construction considers the Baldwin effect with the concept of niche construction; in a very crude form, it states that learning creates a new niche on which both the Baldwin expediting and canalizing effects take place. This simplistic combination of two evolutionary processes provides a mighty explanatory power. However, a close observation will reveal that BNC is not a mere fusion of the two distinctive processes of evolution; rather they synergistically interact and bear interesting properties which are found neither in niche construction nor the Baldwin effect *per se*. As such, they do not bear any priority argument such as host-parasite nor chicken-and-egg type arguments regarding the relationship between niche construction and the Baldwin effect. Although the property of BNC obviously retains the characteristic aspects of both processes, this dualistic property does not easily allow us to decompose it into the two original domains; its property is certainly dualistic but it is rather like the property of light –it is a wave, but a bundle of particles at the same time. Thus, in contrast to other Baldwinian accounts (*i.e.*, Baldwin's breathing-space model and the G-P correlation model) in which particular evolutionary mechanisms serve as a 'framework' for the specific types of Baldwin effect, in BNC, the effect and niche constructions are mutually dependent, or more properly, interwoven with each other.

As a consequence, various new features are found in the theory. These features are not able to be observed in the standard theories of the effect, and as such, they make it theoretically rich and equip the theory with high explanatory power. In this chapter, the property of BNC and some of its implications are investigated.

6.1 Internal Niche Construction in BNC

As the crucial parts of BNC conceptually adopt niche construction, it is assumed that different modes of niche construction may non-trivially affect BNC in different ways. For example, if learning involves physical modifications of environment (*i.e.*, a type of external niche construction), it may take some time for such a niche construction to provide feedback to the organisms. On the other side of the coin, it

also implies that such a modification may persist over generations (*i.e.*, an evolutionary momentum). Although there might be some exceptional cases, it would be generally true that the total amount of physical modifications which are required to change the extant selective pressures may be quite high compared to the magnitude of the possible impacts each individual's activity can make. If so, the pace of BNC will be slow. Especially, if the pace of modification is significantly slower than the species' reproductive cycle, the mode of the Baldwin effect might not be much different from the standard Baldwin effect.

On the other hand, in internal niche construction, the pace of construction is expected to be much faster. The reason is two-fold, although both reasons are fundamentally rooted to the sole interiority property of internal niche construction. The first is no-commitment of physical modification. This is different from external niche construction, given that environmental factors relating to internal niche construction are basically social ones. Therefore, physical environmental modifications are essentially not involved. As such, possible feedbacks of niche construction may quickly come into effect; it is perfectly conceivable in some cases that those individuals who exercise some activity would receive feedback in their own lifetime.

The second point is found in the network-dependent aspect of internal niche construction. In external niche construction, individuals are not necessarily related to each other regarding the given behavior; if individuals' activities modify an environment and it is shared by other members of the same population, there is no need for the individuals to form a social network. In other words, as long as the modified environment is shared in the population, niche construction is conceivable. On the other hand, in internal niche construction, networks are essential; internal niche construction is indeed a process of network creation on which a certain norm/protocol is being shared.

In internal niche construction, not only how many 'successful' learners exist, but also how many of such learners are connected, is an important factor. Connectivity is another way to measure adaptiveness of a certain norm; if a norm creates a large cluster, it would be thought of as highly adaptive compared to other small clusters (*i.e.*, norms) coexisting in the population. Kauffman (1995) provides an excellent insight into this process. Based on the Erdos-Renyi random graph (Erdos & Renyi 1959), he gives an analogical example of "Buttons and Threads"; buttons are connected by threads. When the ratio of the number of threads to the number of buttons in a random graph exceeds little more than 0.5, the size of the largest cluster suddenly jumps up; the connection of nodes (*i.e.*, buttons) become quickly saturated when the number of connections reaches about half of the number of nodes. Given the fact that cultural interactions in a population often exhibit this

type of random network, this 'phase transition' under the low connectivity gives a significant implication on the study of social internal niche construction; in contrast to external niche construction, where the pace of niche construction is usually constant[1], in the case of internal niche construction, it is expected to increase exponentially at some point (most likely an early stage) of the process; it may even become a new STE. This strongly indicates that internal niche construction is a highly 'efficient' mode of niche construction.

In BNC, networks will also affect learning. The more a particular norm becomes dominant (*i.e.*, a network is growing), the more the possibility that learners are exposed to such a norm as the source of their learning. This consequently enhances such a norm to be even more dominant in the next generation.

6.2 The Dual Role of Learning

One of the apparent differences of BNC from other Baldwinian mechanisms is found in the dual role of learning. The first role of learning relates to innovations. As Lewontin and others conceive, a part of the environment is spontaneously created through the organisms' own activity. This is also applicable to BNC; the initial stage of the Baldwin effect could be triggered by innovation which is rooted in the ability of learning itself (*i.e.*, inceptive niche construction) in lieu of autonomous environmental changes so as to emerge as a new selective pressure (*i.e.*, counteractive niche construction). Recall the conventional Baldwin effect (*i.e.*, the G-P correlation model) where the external environment has to change so that it generates a new type of selective pressure independent from a given trait. Under the new environment, learnable individuals become adaptive so that non-learners are winnowed away. On the other hand, in the BNC model, learning itself may create or modify the environment so that it produces a new selective pressure through the niche construction process[2]. In BNC, therefore, organisms are the 'agent' of environmental changes. This is especially true for internal niche construction. In internal niche construction, innovative activities (attributed to learning) can easily and quickly produce new types of environmental conditions. This ease of spontaneous niche construction through learning indicates that this type of the Baldwin effect would be indeed popular especially in species which form social structures based on somewhat ritual behavioral traits.

[1] Although this depends on the susceptibility of a given environment.

[2] Of course, the emergence of the new learning could be the result of a reaction to a new environmental factor. Having said that, it is also perfectly possible that such learning creates an innovation through which a niche construction process spontaneously begins.

To be the igniter of niche construction, a given learning capability has to be available in the initial population beforehand. This assumption may sound somewhat awkward. However, it is perfectly conceivable that such a learning capability was required to be adaptive in the extant environment and already saturated in the population; niche construction is a consequence of its derivational application. Consider, for example, a following case; suppose by using an extant learning ability, a new type of behavior is innovated. If this behavior is useful even a small scale, it would have chance to prevail in a population. Initially, such new behavior may slowly propagate, especially because capabilities of learning typically vary in different age groups; for first few generations only youngsters may well learn such behavior. The propagation of the yam-washing behavior by Japanese macaque (*see* Section 3.1.4) is an example of this. Another example is found in Fisher & Hinde (1949). They reported that in the 1920's, a population of British Blue tits in a village learnt how to open milk bottles to get the cream on the surface by cleverly pecking bottle tops. Village by village, this habit quickly spread[3]. These examples show that for those cognitively sophisticated animals can easily extend their an already extant learning capability to innovate and acquire a new type of behavior.

In these cases, the reported behavior is not social, but rather individual based. On the contrary, if such behavior is socially cooperative, its utility (*i.e.*, adaptiveness) would be expected to increase somewhat proportionally to the size of successful learners (*i.e.*, more and more individuals join to the cooperative behavior). Then, individuals are 'socially trenched' (Deacon 1997). Since the adaptability is mostly determined by the number of individuals who adopt the given behavior, even if the initial innovation is trivial, its impact in later generations would not be negligible. This assures that even a trivial innovation which is a small leap from an extant behavior could mark a major impact on later generations. For example, it would not be so difficult for apes to develop some ritual behavior from, say, grooming behavior so as to avoid unnecessary confrontations. Such behavior would be innovated by a youngster and only shared by its family members initially. However, once such behavior leaps out to other members, its utility will increase dramatically. This frequency-dependent aspect of niche construction will be discussed in the next section. Also, it is important to note that, as described later, learning does not have to be genetically related to innate predisposition (*i.e.*, no need for G-P correlation).

The second role is a conventional one; as a type of the Baldwin effect, learning makes organisms capable of adapting to the new niche. This is somewhat tautological since such a niche is constructed through the learning itself. However, in

[3]Unfortunately this was soon halted as the packaging was changed around the time of World War Ⅱ.

the case of an internal niche, this becomes quite an important factor; once such a niche is created within a fraction of the population, other members are better able to align with the norm to be adaptive. To align, learning plays a vital role.

In summary, in Baldwinian niche construction, learning plays two crucial roles in the early stage of the process. Once such a niche is constructed, however, learning might become redundant; the cost created by learning is disfavored by selection and consequently the canalization process takes place. This process is basically the same as the conventional Baldwin canalizing effect.

6.3 Positive Frequency-Dependent Selection

As noted in the above section, another important aspect of BNC is that it is a primarily positive frequency-dependent selection. Positive frequency-dependent selection is a mode of natural selection which favors a common phenotype in a population. This tendency is again salient in internal niche construction. When a norm is established and yet shared by a small fragment of a population, unsuccessful learners (regarding the particular norm) are still fairly adaptive. This is due to such unsuccessful learners being thought of as already adaptive to the extant environment. Therefore, the initial dispersion of such a norm would be slow. However, as the successful learners enjoy their high fitness among themselves, eventually the number of learners increases as a consequence of the winnowing process of natural selection. Once this upward process starts, positive frequency-dependent selection provides a dynamic evolutionary trajectory; it typically shows a sigmoidal curve when one plots the number of adaptive members. In the case of BNC, this corresponds to the number of (successful) learners.

In the case of cooperative niche construction, frequency-dependent selection tends to lead to an equilibrium rapidly. That is, as the evolutionary process proceeds, penalties against non-learners or unsuccessful learners quickly become very severe. Therefore, at the end of the selection process, almost every member of the population exercises the same learnt behavior. This equilibrium, however, has a space to be modified; if learning is costly, by canalizing it one can increase his fitness without breaking (*i.e.*, maintaining) a given niche.

This frequency-dependent aspect of BNC shows an intriguing insight for the study of BNC. Interestingly, Arthur's (1994) study in economics would be very informative. In conventional economics, it is almost dogmatically accepted that the value of a given good decreases if availability of the item increases in a market. This is called "*decreasing return*". What Arthur argues is that certain goods whose utility depends on a network do not follow this rule; he has convincingly shown that in the economical market a certain range of quality and cost differences are

ignorable if some of a given good's value is determined within a network, but not by the item itself. Under this circumstance, values of such goods actually *increase* as the number of users increase. This is termed "*increasing return*".

A typical example is found in "the VCR war"; it is a well-known fact that *BETA* type was better than *VHS* regarding its quality. Also *VHS* and *BETA* were introduced at roughly the same time for roughly the same price. Initially, the market was unstable. However, since there was no compatibility between these two systems, as more *VHS* VCRs were purchased, video stores were encouraged to stock more *VHS* tapes. This consequently increased the value of the *VHS* players and therefore more *VHS* players were purchased. Once the trend was established, the whole VCR market was firmly entrenched in the *VHS* system. As a result, *BETA* was completely washed out from the market. In this case, the difference in quality of the two VCRs was almost irrelevant to the final consequence.

This is exactly a case of positive frequency-dependent selection; by a positive feedback process, utility (or value) of a particular good increases as the number of the goods' user increases. Note that the key factor of this frequency-dependent selection is 'compatibility' of the two systems. Like most other electrical appliances, VCR does not fundamentally require network systems; both *VHS* and *BETA* work perfectly on their own. However, as people started exchanging their contents through their complementary goods (*i.e.*, via video tapes), the network aspect (through the compatibility) became important; consequently selection takes place. Since in the case of internal niche construction, conformity of norms are the factor of this network aspect, a similar phenomenon would be perfectly conceivable.

6.4 Network Externality and Consistency over Contingencies

Arthur considers that the reason a specific good is chosen is mostly accidental, if differences of goods' qualities are not large. This is well reflected in the above case; quality-wise, *BETA* had a slightly better advantage. This shows that actual utility of a specific good is determined within a given network but not on its own. This is called "*Network Externality*". As exemplified above, internal niche construction fundamentally involves this aspect.

The network externality aspect of internal niche construction provides some interesting properties. Positive-frequency dependence is one such property. Secondly, in contrast to other types of the Baldwin effect, in Baldwinian internal niche construction, a peculiar contingency exists. Suppose that two different populations are prepared from the same strain. Each population is isolated from the other. When they are let evolve, suppose also that they start developing a communication system; internal niche constructions take place. Although each population may create

an idiosyncratic communication system (*i.e.*, niche), from a metaperspective, they are most likely qualitatively equivalent. In other words, while actual protocols, styles, or manners would be different, overall properties would be the same or very similar. In other words, it is the overall quality of such niches that is consistent –"*consistency over contingencies*".

An even simpler example will help the understanding of this property; imagine the spontaneous emergence of a traffic lane system (*i.e.*, the right- or left-hand side systems) in a region where previously no rule was installed. After numerous collisions perhaps, the system will quickly settle into either the right- or left-hand side system. Once such a system becomes settled, no obvious qualitative difference exists; it is somewhat meaningless to argue the superiority of one traffic system over the other.

Although the cases are very simple and only allow the two different systems, they accurately point out the pertinent part of the property. Indeed, the traffic system can be conceived of as a non-evolutionary internal niche construction. While no genetic evolution is involved, establishing a specific traffic system is thought of as a norm creation process; locally interacting agents are spontaneously establishing an order through creating conformity.

The origin of such a consistency would be a result of combination between some external pressures which are ubiquitously available for different populations and population internal factors which collectively create a complex adaptive response to such pressures. Regarding behavioral/psychological properties, as Bates *et al.* (1998) discuss, such external pressures will be found in STEs and human cognitive capacity; cognitive processing problems (like language-mapping problem) often require experiences which are shared by all normal members of a species. Such experiences are an STE and they induce some ineluctable responses. It is this ineluctability that is the source of consistency created by complex adaptive responses as described below. Such an STE is created through a niche construction process, and gradually shared by the population. The external pressures also make the channels used for such processing be subject to universal constraints.

On the other hand, complex adaptive responses are often conceived of as emergent phenomena created by local interactions of constituents of a population. Generally this type of emergent phenomena involves stochastic factors; as complex adaptive responses are collective and synergistic interactions of local constituents, which are not globally controlled, accidental factors easily penetrate. However, as an adaptive system, the populations respond to the external pressures by finding a good solution.

Thus, interestingly, in BNC which involves some network external aspects, an STE produced through the niche construction process serves as a continuous, stable external pressure for complex adaptive responses. This suggests that increase of ineluctability through BNC is a synergistic reaction of STE and the canalizing effect. This point will be briefly discussed in Chapter 10.

This network external aspect also brings a "*lock-in*" effect (Arthur 1994) to the complex adaptive system. In the case of economy, once a trend is set, because of the frequency-dependent aspect of the goods, users cannot move to other goods without substantial costs. This lock-in effect is conceivable as the reason for contingencies. Also this type of system is sensitive to previous generations' conditions which are inherited via both genetic and extragenetic inheritances, such a system has the "*path-dependent*" aspect because accidental events might have a persistent effect on its course. This is another term for the system is "dynamic".

What makes these properties especially intriguing for the study of BNC is that the possibility that what the canalizing effect can incorporate could be not individual instances of niche constructions, but the part of the consistency across different niches[4]. As described below, if evolutionary momenta work at this point, some cumulative universalities may emerge in an evolved system. However, this is beyond the range of our concern in this thesis, so we do not touch on it any further here.

6.5 No Mutation, No Correlation

As a type of niche construction, exaptation processes through bilateral feedbacks would be the primary engine of genetic evolution. This is one of the most crucial aspects of BNC. This implies that, because of the exaptation process, genetic operations are not essential. That is, learning itself creates a new type of selective pressure, previously unexposed differences in genotypes (by natural selection) are now exposed to the new winnowing process (*i.e.*, exaptation). This indicates that along with the increase of successful learners, some individuals may have already become 'naturally' more ineluctable; their genotypes become innately adaptive under the new environment without any modification.

In this case, therefore, it is assumed that the canalization process in BNC is most likely powered not by mutations, but rather by the diversity of the genepool, although BNC does not exclude the possibility of mutation-based evolutionary scenarios. Note that Waddington himself also stressed this point in the concept of genetic assimilation (*see* Chapter 2). The most significant property of BNC derived from these aspects will be, however, about G-P decorrelation. That is, phenocopies are not necessarily genetically close to their corresponding ineluctable phenotype.

[4]This is remotely related to Avital & Jablonka's (2000) categorizing effect.

Waddington's genetic assimilation not only states that the innately predisposed trait takes over the learnt traits, but also literally assumes that those phenocopies 'genetically' assimilate to the preformed trait; better learners' genotypes have to be genetically close to the innately predisposed genotype. Recall that this is the reason that we think that the term "genetic assimilation" entails the two different levels of phenomena; the assimilation of genotypes (*i.e.*, the genetic level), and the assimilation of learnt behavior (*i.e.*, the phenotypic level) –assimilated by a *genetically* predisposed trait. However, in BNC, this requisition of the genotype-phenotype correlation is redundant. As long as learners create a new niche which is sufficiently accessible for other members, then a genetically distant, yet functionally similar innately predisposed trait can be selected out.

It is no exaggeration that this causes a profound impact on the study of the Baldwin effect. With this property learning can be a more domain general capacity whose genetic basis is not specifically related to a given innate predisposition. Since in the G-P model of the Baldwin effect, both learning and innate predisposition are genetically related, it is natural to conceive that the ability of learning is highly domain specific. As a domain general capacity, learning permits a broader range of evolution than what the G-P correlation mechanism can think of. After all, exaptation is the cheapest evolutionary trick evolution can hire. This consequently broadens the chance that such learning leads to BNC processes. Also, this type of general learning ability would be equated to general intelligence. This implies that higher organisms would have more chance of being involved in BNC processes. These are other reasons that BNC may be much more general than other mechanisms of the Baldwin effect.

In BNC, the expediting effect exhibits an idiosyncratic process; in the conventional mechanisms, learning increases the pace of the evolutionary process because it fills the gap that evolutionary search cannot bridge. Indeed, in both Baldwin's breathing-space model and the G-P correlation model, learning contributes to reproduction so that in such a proliferated population, more ineluctable individuals are expected to emerge due to genetic operations[5]. Thus, although the degree of expedition would differ between these two mechanisms, fundamentally they share the basic mechanism of the expediting effect. On the other hand, the genetic-variance-driven evolutionary process of BNC is indeed a very significant point which differentiates BNC from other Baldwinian theories; niche construction serves as an 'unmasking' process (Deacon 2003) on the extant genetic differences among the individuals in

[5]Recall that the difference between the breathing-space model and the G-P correlation models is that while the breathing-space model did not consider genetic similarity between successful learners and more canalized individuals, the G-P correlation model stands on this assumption.

the population[6]. Since the required genetic variance would already be available, mutations are redundant. This greatly enhances a pace of evolutionary process. Godfrey-Smith (2003) has seemed to miss this point, although Dor & Jablonka (2000, 2001) concisely describe it.

The greatest benefit of a rapid evolutionary process powered by the 'unmasking' (*i.e.*, exaptation) effect is socio-cultural evolution. Given the consideration that socio-cultural evolutionary processes would be far faster than genetic evolution, the redundancy of mutational evolution would be highly appealing. Thus, together with the frequency-dependent selection, BNC creates a very rapid evolutionary process. Although somewhat reminiscent of Waddington's concept of genetic assimilation, this indicates that the diversity of the initial genepool is a crucial factor in BNC.

There is another crucial aspect in this genetic-variance-driven mechanism of the Baldwin effect. That is, canalization now does not wait for the completion of Stage 2 of the Baldwin effect; both in Baldwin's breathing space model and Hinton & Nowlan's type of G-P correlation model, the saturation (or near saturation) of learners in Stage 2 is more or less required. This bridges the non-adaptive populations (Stage 1) and the innately adaptive populations (Stage 3). However, in BNC, Stage 2 is significantly truncated; if both learning and mating locally take place, full saturation of successful learners in the entire population is unnecessary. Then genotypes which can be already recognized as innately adaptive in the context (i.e, in the given niche), if any, will be selected within such a local domain. This consequently encourages development of 'dialects' in the population.

6.6 Cycles of Exaptation & Canalization and The Assimilate-Stretch Principle

Together with no need of G-P correlation, it is perfectly plausible to consider that learning is a domain general ability; it can be used for acquisitions of different types of abilities. If this is indeed the case, Avital & Jablonka's (2000) assimilate-stretch principle can be applicable (*see* Section 2.7). Recall that Avital & Jablonka consider that the domain to which this principle can be applied is sequential learning. There are two conceivable reasons for this. First, the ability of learning should be functionally constrained. That is, as a type of G-P correlation model, Avital & Jablonka have to stipulate that the genetic attribution of learning has to be closely aligned to more ineluctable, more innately predisposed yet slightly different behaviors. If such behaviors have to be genetically close each other so that learning can be

[6]It is somewhat sarcastic since learning is usually considered as "masking" genetic differences of individuals, in BNC, it works as the ultimate source of exposing such differences. *see* Chapter 10 for a further discussion

continuously canalized to such behaviors. Under this stipulation, it is more natural to consider that such behaviors form a sequence.

Secondly, in Avital & Jablonka's model, environmental conditions are thought of autonomously changing from organisms. This is because their model does not consider niche construction (*i.e.*, the G-P correlation model). Then for assimilate-stretch to take place, different behaviors, *as a whole*, have to have the same solution to a specific environmental pressure. In this case, sequential behavior is a natural assumption.

On the other hand, BNC does not limit the assimilate-stretch principle to be exclusively applied to sequential learning. Apparently this comes from the fact that organisms are capable of modifying environmental conditions subjectively. Once a canalization process is complete, learning is free to create another niche. That is, learning is now ready for creating a new norm in the population. It might be a totally different type of niche from the previous niche.

However, what is interesting is the case that a new niche is created on top of the previously created niche. In other words, learning elaborates the extant norm so that a further sophistication of the norm itself follows. Then yet another canalization may take place. If these processes take place in a cyclic manner, within a relatively short period, a highly sophisticated norm (or a system of norms) may emerge. This is somewhat similar to Avital & Jablonka's original formulation of the assimilate-stretch (*i.e.*, to sequential learnings), the degree of freedom in this scenario is much higher.

6.7 Open and Long Causal Chain

Odling-Smee *et al.* (2003) stress that the feedback process of niche construction is not confined to a specific set of loci which is responsible for the niche creation trait. Rather what the trait creates would trigger evolution of otherwise causally irrelevant traits; a newly created niche exposes such traits to different types of selective pressures. This open-ended causal chain enables us to consider a cascading effect on the evolutionary process. Consequently, possible causal links may be long and as a result, it may allow other factors to intervene in the causal chain. The curious causal relation between yam cultivation and high frequency of sickle cell anemia in West Africa is a good example of this. Through the yam cultivation lands are opened, consequently some previously unexposed ponds become rich beds for fertilizing mosquitoes in genus *Anopheles*. This leads to high malarial contaminations in human population. High frequency of sickle cell anemia is the adaptive response to this contamination; the allele responsible for the anemia is also deeply involved

in malarial resistance, if it appears in a heterozygote form. Thus, balancing selection has taken place for this double-edged gene. However, it is also true that prevailing malaria in a given population is sensitive to the amount of mosquitoes or the size of the pond, and so on. Therefore, some uncontrollable factors serve as interventions to the already somehow overstretched causal chain[7]. Although it is fascinating to consider linking the evolution of two or more distant traits, it is obviously a double-edged expansion of the niche construction theory. While it enjoys a strong explanatory power, it may jeopardize a clear causal relationship of a given evolutionary event in a messy state[8].

One apparent reason is that such studies incorporate 'functionally irrelevant' traits (of the niche-creating trait) in their research range; this inflates research targets, and consequently leads to long sloppy causal chains. Fortunately, BNC, in principle, deals with functionally similar or related traits in any given study; although BNC does not confine loci of learnt or innately predisposed behavior (*i.e.*, at the genetic level), it limits the functional aspect (*i.e.*, at the phenotypic level).

6.8 Dual Inheritance and Evolutionary/Learning Momenta

Recall that extragenetic inheritance is one of the key properties of niche construction. As described in Section 5.2 evolutionary momenta emerge when this extragenetic inheritance takes place. This is especially likely in the case of external niche construction where physical environmental conditions easily remain longer than one's life. However, even in a case of internal niche construction, a created norm is often inherited to next generation. Since BNC involves learning, this type of inherited social norm is naturally assumed.

Moreover, since BNC involves the canalization process, a different type of momentum is conceivable. Through the canalizing effect, at least some part of the norm becomes supported by an innately predisposed, highly ineluctable trait. Then, what becomes highly ineluctable through the canalization process becomes a basic constraint of the next niche construction process. In other words, what previously canalized through BNC becomes a foundation of future norms. If, as described above, BNC takes place cyclically, this mode of genetic inheritance brings a very

[7]Indeed, Odling-Smee *et al.* consider that the theory of niche construction is, exactly at this point, different from the so-called *gene-culture coevolutionary theory* (Boyd & Richerson 1985).

[8]To provide an explanation for some evolutionary events, descriptions of such long causal chains may be necessary. However, as it allows a number of both evolutionary and non-evolutionary factors to be involved (*e.g.*, interventions of abiotic or conspecific factors), the chance of establishing rigorous causal explanations would become slim. In other words, it brings about another 'just-so' story in evolutionary study.

interesting scenario. Furthermore, if what is canalized is properties which consistently appear in different niche constructions (*i.e.*, consistency over contingencies), what cyclic BNC can bring is universal constraints of a new niche construction.

At least two significant implications are derived from this observation. First, BNC enables organisms to inherit some portion of a previously culturally-inherited trait in a much longer term without direct inheritance of the target trait itself. This is saliently different from the majority of cultural evolution where discontinuity of such inheritances means immediate liquidation. On the other hand, by bringing the organism-internal inheritance mechanism, BNC leaves the option that two temporally segregated (and consequently culturally separated), but genetically connected populations create similar niches. Of course, this type of momentum may well produce temporal maladaptations in the current environment. Note, however, that as a case of internal niche construction, the environment which defines organisms' adaptiveness itself is created through the niche construction process which is constrained by the organism-internal factors. In this regard, severe maladaptations are avoidable from the beginning.

Chapter 7

Baldwin Niche Construction and Language Evolution

Given the described properties of BNC, in this chapter, the reasons that BNC is the appropriate mechanism of language evolution are examined.

7.1 Theoretical Inadequacy of the Conventional Baldwin Effect

As in the literature review, non-trivial numbers of researchers pay attention to the explanatory possibility of the Baldwin effect. However, unfortunately almost all of such researchers fail to recognize the conventional Baldwin effect is fundamentally inappropriate to be applied to language evolution.

The most serious problem is the discrepancy between the theoretically somewhat optimistic assumptions of the conventional Baldwin effect models and the complexity of the linguistic ability. The majority of the language evolution studies which concern the Baldwin effect assume some form of 'genetic assimilation' (canalization) as a core mechanism of the effect. Then the pertinent question to be asked is how sensitive the necessary conditions of the genetic assimilation model are in the context of language evolution. Recall that, in the Waddingtonian model, G-P correlation is the key prerequisite for the model to work –good learners are more innately predisposed. Mayley, Yamauchi and Briscoe convincingly demonstrate it. However, a theoretical concern has been haunting this simplistic assumption even outside of the domain of language evolution. That is, for higher order phenotypic traits such as behavioral, psychological, cognitive abilities, this assumption of genetic linearity seems to be too naïve.

Recall the apparent, but often neglected fact that all genes can produce is simply a set of amino acids. Given this, it is unmistakably clear that any behavioral and other higher order traits cannot be directly manifested entities of genes; there is a causally large leap from simple amino acids to, say, psychological traits. Although this does not immediately dismiss the idea that the genes 'control' such higher order phenotypic traits, it certainly raises an alarm for such an optimistic assumption.

One shall not forget that this is one of the fundamental motivations of Waddington's central field of epigenesis. Also it is the same driving force that leads Lewontin to consider the constructivist approach where such a leap is filled by interactions between genotype and environment during development.

While the importance of epigenetic development and genotype-environment interactions does not directly jeopardize the assumption of G-P correlation, these shed important light on the complexity of possible G-P relationships. That is, the straightforward relationship between genotype and phenotype is rather a special case but not dominant at least at the level of higher order traits. Genetically different, but functionally similar traits may be involved in a given phenotypic domain. In the case of the Baldwin effect, three different levels of decorrelation between genotypes and phenotypes are conceivable.

The first type is that the genetic foundation of learning is completely decorrelated from that of the innate predisposition. As described in Section 6.5, this is a perfectly possible scenario in BNC, but is devastating for the conventional Baldwin effect; any genetic operation at the loci responsible for learning may not affect the loci for the innate predisposition[1].

Despite this difficulty for the conventional Baldwin effect, learning as a domain-general cognitive capacity is strongly appealing for the study of language evolution. Applying such a domain-general learning to acquisition of (primitive) linguistic knowledge is a type of exaptation. Furthermore, in contrast to other scenarios, this complete decorrelation model accepts cycles of canalization processes described in Section 6.5. Since the possibility of cyclic, assimilate-stretch process in the context of the complete G-P decorrelation has been already described, we do not repeat the explanation here. However, it is important to note that BNC is the only conceivable mechanism (and positively utilize the condition) in the domain of domain-general learning regarding this process.

The second case is that though genetic bases of both learning ability and innate predisposition share the same loci in a genotype, the target (*i.e.*, optimal) genotypes are different. This is basically the same as the argument described in Section 4.2.2 and Best's study (*see* Section 2.8.5). The situation is especially destructive, if such target genotypes are negatively correlated (*i.e.*, increasing the frequency of genotypes for learning work for decreasing the frequency of the innate predisposition). In this situation, the harder selection attempts to optimize learning, the smaller the chance that the population has the innately predisposed individual. This is

[1] An exception is if the locations of such loci are close to each other so that some *hitchhiking effects* (Maynard Smith & Haigh 1974) are expected. However, as shown below, this poses a different problem.

also problematic for BNC, since both learners and the innately predisposed are advantageous in a new niche. Consequently there would be a 'pulling game' between genotypes for learning genotypes and that of innately predisposed individuals in a population. However, in contrast to the conventional Baldwin effect, the situation is not so catastrophic; as a niche being ubiquitously accessible (by learners), the innately predisposed individuals also become adaptive. Therefore, as long as the genepool holds a sufficient diversity, there is a great chance that the frequency of the innately adaptive genotype increases.

The third case is about genetic interactions. Recall polygenic inheritance; epistasis and pleiotropy are two closely-related properties of polygenic inheritance –the former states that two or more genes non-additively contribute to one trait, while the latter describes one gene that contributes to two or more separate traits.

When polygenic inheritance exists, the relationship between a genotype and its manifested entity qua phenotype is disturbed. That is, if the relationship between the genotype and the phenotype is epistatic, these are, to some extent, decorrelated. A simple explanation provides a clear view for this. Consider a given allele's effect on a phenotypic expression. If there is no epistasis, then substitution of the allele to other alleles will be directly reflected on the phenotype. In other words, the difference between the two alleles on the phenotypic level is simply measured by the difference between the two phenotypes. Basically the same thing can be applicable to additive polygenic inheritance; if a specific phenotypic trait exhibits an additive effect, the difference of two different alleles on same locus can be measured by the difference of two corresponding phenotypes as long as other genes remain the same. However, when such polygenic effects becomes non-additive, the situation becomes different. Since the given phenotypic trait is sensitive to different alleles in two or more different loci, and substitution of such alleles is non-linear, a possible effect of a replacement of a given allele would be idiosyncratic if other loci remain the same. In other words, there is no general direction in the phenotypic level when the genetic changes are directional. Thus, it is apparent that it is a case of G-P decorrelation. It is assumed that as the number of genes involved in epistasis increases, the degree of decorrelation also generally augments.

Note that those three cases are nothing special. Given the fact that linguistic competence is the most complex cognitive ability living creatures ever have, it is totally natural to assume that the genetic basis of linguistic acquisition ability (*i.e.*, learning) is completely different from that of innate linguistic knowledge (*i.e.*, innate predisposition), although they are functionally similar. While this is a case of complete decorrelation of learning abilities and innate predispositions, others are equally conceivable. For example, consider the case of epistasis. Recently,

a growing body of research accepts the fact that epistatic interactions are fairly common in genetics and, the popularity of simple additive assumptions is merely because of theoretical requirement but not based on empirical reasons. In other words, these types of genetic interactions in phenotypic expression (*i.e.*, epistasis) are considered to be quite common; indeed it is conceived that simple additive cases are the 'exceptions' (Rice 2000). Therefore, it is easily assumed that, as the most complex cognitive ability, the genetic foundation of linguistic knowledge involves a high degree of epistatic effect. Given this, the application of the G-P correlation model of the Baldwin effect (and, needless to say, Baldwin's original model) becomes progressively dubious.

On the other hand, BNC is remarkably robust in such genetically 'tough' conditions; the case of the complete decorrelation gives a good idea of this; through the combination of niche construction and exaptation, two genetically independent phenotypic traits can interact. The other conditions are also surmountable. Consider, for example, the case of epistasis. While epistasis disturbs the genetic level of 'assimilation' to be reflected at the phenotypic level (and this is deadly problematic for the G-P correlation model), BNC is not largely affected by this disturbance; for this type of decorrelation to be effective, a process has to involve some genetic operations. However, as described in Section 6.5, such genetic operations are not essential for BNC to perform. Therefore, the epistatic decorrelation is assumed to be largely negligible in BNC.

7.2 Theoretical Necessity of BNC

In the above section, we considered the theoretical plausibility of BNC in language evolution by eliminating other possibilities. In this section, instead, the intriguing relationship of BNC and language evolution is briefly discussed.

When language is captured as a system of communication protocol, 'conformity' or 'parity' would be the most closely related aspect to internal niche construction. Pinker & Bloom (1990) nicely described this aspect. Linguistic communicators in a given population have to share some common coding protocol (Liberman & Mattingly 1989). However, such a protocol does not have to be shared with other populations (*i.e.*, network-externality). In other words, an arbitrary coding protocol idiosyncratic to a specific population is functional as long as it is shared by members in the population. At different levels of linguistic communication system, this type of protocol parity would play important roles.

Waddington considered this point from a more linguistic point of view and termed it "convention" in his essay on language evolution (*see* Section 3.1.1). This

is exactly equivalent to our concept of norm in internal niche construction. The following points are shared by both parities and norms:

1. Conformity: Only by following such a parity or norm, can individuals gain benefit. As conformity is based on cooperative behavior of individuals, cooperative niche construction is most related.
2. Network Externality: Both parity (convention) and norm are only population internally meaningful; in the same manner as parity, a created norm through niche construction is only meaningful among members in a given population.
3. Sociality: A system that consists of such parities or norms only exists in a social network.

Together with the discussions in Chapter 6, these similarities between linguistic parity and niche construction ensure that BNC is the most plausible mechanism of the Baldwin effect in language evolution. Somewhat surprisingly, therefore, Waddington had already suggested that language evolution is indeed a case of internal niche construction. It is a pity, then, that Waddington, as the inventor of the modern mechanism of the Baldwin effect, did not find that there is a totally different mode of canalization in niche construction –BNC.

Chapter 8

Linguistic and Genetic Representations

So far, we have developed the theory of BNC based on the critical review of the conventional Baldwin effect. In summary, while it is thought of as the existence of a complex relationship between learning and canalized linguistic knowledge, the mechanism of the conventional Baldwin effect only admits a fairly simple relationship. In other words, the G-P correlation model only inadequately captures a possible trajectory of language evolution. On the other hand, because of the exaptation process triggered by niche construction, BNC will robustly take place in decorrelated circumstances in genotype-phenotype mappings. In the rest of this thesis, we will examine this with computer simulations.

To model BNC in language evolution in computer simulations, we have to consider both genetic and linguistic representations. Although there are various representations are conceivable both in genetic and linguistic levels from very simple ones to highly elaborated, complex ones. Given that the main purpose of this thesis is to propose a new type of mechanism which constitutes a vital part of the Baldwinian account of language evolution, we adopt simple representations which can provide clear causalities are more suitable.

To attain this, we basically adopt our model of language to Turkel (2002) and Kirby & Hurford (1997). Recall that based on Hinton & Nowlan's model of the Baldwin effect (Hinton & Nowlan 1987), Turkel brings a popular concept of the LAD in generative linguistics, namely the P&P approach (*see* Section 3.2.1).

Adopting a particular theory of linguistics (thus particular linguistic representations in both genetic and cognitive levels) does not necessarily reflect our belief about language at these levels. Rather, the whole motivation behind of this application of the P&P approach is somewhat parsimonious in several senses; first, the approach is arguably the most well-studied theory in language acquisition. It has been a core framework of generative linguistics, the most popular camp of linguistics. Secondly, this is almost the only approach which encompasses not only general

linguistic phenomena but also language acquisition with some (even though very crude) considerations of genetic foundations.

However, this does not necessarily mean that adopting such a theory is buying into their view of language (especially language acquisition and its genetic foundations). The P&P approach is a general theoretical perspective in generative linguistics, and a framework for considering issues in language acquisition. Thus it is not a theory in the sense of a specific model that provides some testable predictions. Therefore, even within this approach, it would be possible to take a somewhat neutral stance regarding the nature & nurture debate to some extent, although its theoretical foundation rests on a strongly nativist view. In other words, the approach accepts a some degree of plasticity for its theoretical implementations. Briscoe's implementation of principles/parameters in his GCG-based simulations strikes this point. This is the third point. Finally, this approach exhibits a strong affinity with computer simulations. This point is well represented in the studies of Turkel and Kirby & Hurford compared with their counterpart, namely Hinton & Nowlan. Therefore, practically by hiring the P&P theory, we can construct a model which maintains consistency across different simulations (*i.e.*, Hinton & Nowlan, Turkel, Kirby & Hurford, and simulations in this thesis). This contributes for us to having a good perspective, as it allows to make clear comparisons with both linguistic and non-linguistic Baldwinian simulations.

8.1 The Principles and Parameters Approach

Because our simulations including replications of some previously presented studies adopt the P&P based representations, it is a good idea to review the basic concept of the approach. In this section, a brief explanation of the theory is presented.

As discussed in Chapter 1, despite the considerable differences of surface structures of various languages, it can be agreed that all natural languages are equally complex; indeed the most complex system in any cognitive faculty. For example, for some languages, a predicate-agent relationship is expressed by rich morphological cues, while in some languages, by strictly fixed word orders; for others it might be a mixture of the two. The question arises, then, how any child, wherever he may be in the world, can ineluctably acquire such a complicated system. In other words, language acquisition is a capability unique to humans as a whole. Everyone intuitively knows linguistic input plays a crucial role in language acquisition, as the story of Genie vividly tells (Curtiss 1977). The story is surely an extreme case but it reveals how crucial linguistic input is in language acquisition.

Somewhat contradictorily, however, such vital linguistic input employed to construct knowledge of a language is often ill-formed, incoherent, and most importantly,

insufficient (Chomsky 1965). In other words, compared to second language learners, children do not seem to get sufficient qualitative and quantitative linguistic instructions from adult speakers[1]. One of the clearest examples of this insufficiency is the absence of so-called "negative evidence". As a part of the insufficiency, children do not get regular feedbacks for their grammatical mistakes, while such information is vital for any second language learners. Consider, for instance, a case of a poor second language learner who cannot have access to any negative feedback; it is almost impossible for her to construct a proper knowledge of the second language without accessing such information especially after puberty (post critical period). If language acquisition is a matter of learning, there is no explanation for the fact that first and second language learners' language proficiencies are markedly different. Thus, claims have been made that the process of language acquisition is neither completely a process of learning nor a product of fully canalized linguistic knowledge.

Note, however, that although this is a reasonable conclusion, the claims themselves do not provide any answer to the problems stated above. After all, any complex cognitive abilities are complex products of both innately prespecified knowledge and learning. Even a young chick has to be equipped with both innately prespecified knowledge and post-natal experiences to complete its attachment to its mother.

The problem is what innate knowledge and post-natal language acquisition look like. This is an extremely complicated question. For example, language acquisition is not so amenable to a simple deductive algorithm. Instead, it is widely understood that syntactic rules are often optional (Grimshaw 1981); it is common that a syntactic rule is applicable to some subsets of a syntactic category whereas other subsets of the same category do not allow the rule to apply. A number of researchers report that children often generalize certain grammatical rules over the actual range of the rule's applications. Even in a simple syntactic rule, we can find a number of exceptions. If the exceptions themselves form their own regularity, however, the problem is not so complicated –the algorithmic deduction is still applicable to the sub-regularity itself.

On the other hand, if such exceptions do not form any regularity, the problem becomes more serious. For example, English has quite a productive verb prefix system such as reversative *un-*. Although a fairly simple rule is required to form '*un* + *VERB*', there are certain verbs that refuse to accept this reversative affix. Information that tells children which verbs refuse the affixation is not part of the information present in the affixation rule itself. Rather, the criteria of this exception

[1]However there is a growing body of research that reverses this conventional view (*see* Kirby 2000, Ellefson & Christiansen 2000*b*)

is unclear when a child hypothesizes the rule. Subsequently, she often produces ill-formed reversative verbs. The crucial problem is that the kind of data which is required to check the validity of her hypothesis does not seem to be the same type of data as the kind the child actually receives in any stages of her language acquisition. This is an embarrassing problem for the children if they utilize the algorithmic deduction scheme for language acquisition (Baker 1979).

Consider this point with a metaphor of card magic. Suppose somebody plays a card trick in front of you. If the trick is a simple one, your first impression will fade quickly; with a few repetitions, you may find out what is the basic trick behind the magic by looking at the step-by-step operations (= algorithmic deduction). A truly impressive act, however, violates your assumptions when you expect a certain result of enabling a complete comprehension. At the most perplexing moment, the magician provides a completely unexpected result. All of a sudden, your optimistic guess falls apart and you are stranded in the middle of nowhere. Indeed, this is the fundamental idea of magic; compilation of algorithmic operations producing a totally implausible result. Language acquisition often seems like a big magic trick. Given the fact that we cannot often find out the basic tricks underlying even a little complicated magic, the child language acquisition is a truly striking process.

If you had studied card magic before, however, the situation would be totally different. If the magic were some variation of the magic you learned before, after some trials and errors you might find out what happened behind the magician's hands. Existing knowledge guides our algorithmic deduction. This is the basic idea of Chomsky's (1981) original formulation of the nature of the LAD and its core theory –the Principles and Parameters theory. In the approach, two different types of limited innate linguistic knowledge are available for children –'Principles' and 'Parameters'. Principles are universal among all natural languages and considered as fully prespecified knowledge. Parameters are partially specified knowledge that take binary parametric values. Setting of each parametric value is triggered by post-natal linguistic experiences. Therefore, the difficulty of algorithmic deduction is circumvented by the partially pre-determined information provided. The differences between natural languages are also attributed to their idiosyncratic configurations of parameters.

Let us look at an example of P&P approach. All languages are thought to have certain lexical heads which can assign case to their arguments. However, languages are different in the grammatical categories of the case assigners. Children have to hypothesize a possible grammar of language. In the P&P approach, this is coded in binary form by either:

- Encoding parameters as descriptive statements about the target grammar that may bear a truth value,
  OR
- Fixing an order of the parameters.

The binary codes, 0's and 1's might be interpreted as the presence or absence, or the alternative forms of grammatical rule(s) of the target language. The following example is taken from Clark (1994);

1. Nominative Case is assigned under SPEC-Head agreement with Tense.
2. Nominative Case is assigned under government with Tense.
3. Exceptional Case Marking is possible.
4. Structural Case Marking is possible.
5. Verb-Second is obligatory in the root.

These can be coded as:

- < 1 0 1 0 1 >
  THIS MEANS:
- Nominative Case is assigned under SPEC-Head agreement with Tense but not under government; the language allows ECM but no SCM; V-2 is obligatory in the root.

We can picture the possible mechanism of the LAD as an incomplete learning device where certain binary information is missing. Thus, setting the values of the parameters is equal to working with the missing information.

This approach provides an appealing model of language acquisition that greatly reduces the complications of conventional acquisition models. In summary, it provides an explanation for important aspects of language –universal learnability and language diversity. On the one hand, universal learnability of natural languages is captured by the limited innate linguistic knowledge. All natural languages sit within a certain range which can be specified by principles and types of parameters. On the other hand, the diversity of languages is attained by different settings of parameters. In this regard, the P&P approach has succeeded in connecting language diversity and universal learnability, which had conventionally been considered as a paradox. Typological diversity of language and universal learnability are rooted in the same property, namely partial innate linguistic knowledge. In this approach, the innate linguistic knowledge 'sets' the degree of these two superficially distinct properties.

Chapter 9

The Experiments

In the previous chapters, we have seen an important development of the new concept of the Baldwin effect, namely Baldwinian Niche Construction. This concept is created from the reconsideration of previous mechanisms, especially the G-P correlation model of the Baldwin effect. The mechanism is based on naïve assumptions regarding genotype-phenotype correlation.

In this chapter, several simulations are demonstrated to investigate both susceptibility of G-P decorrelation in the conventional Baldwin effect and feasibility of BNC. First, some replications of simulations are presented. These simulations are designed to demonstrate the Baldwin effect both in standard evolution and language evolution. Then some types of G-P decorrelations are examined based on these replicated simulations. Finally, simulations based on BNC are investigated.

9.1 Replications

9.1.1 Hinton & Nowlan

In this section, the simulation conducted by Hinton & Nowlan is replicated. As the simulation serves as the base of all other simulations shown in this thesis, it is a good idea to spare a section to reexamine the result briefly.

Hinton & Nowlan's model of the Baldwin effect is a typical example of adoption of the G-P correlation model. In the simulation, apart from extremes where genotypes are either completely ineluctable or fully plastic, any genotypes are some mixture of learnable and innately bounded alleles. They are in an allelic relationship, since these potentially occupy the same loci as one or the other. This is one of the most explicit implementations of the G-P correlation model. However, the objective function shows that the model does not have an additive fitness measure; the 'needle-in-a-haystack' fitness landscape is an example of 'epistasis for fitness' (*see* Section 5.2). Note that, as such, this model is not a model of epistasis which we have discussed; in the thesis, epistasis in the sense of genic interactions has been discussed. Recall that this is called "epistasis for phenotype".

In the simulation, to make it comparable to other simulations, the number of genes in a genotype is truncated from 20 (the original configuration) to twelve. The size of the population is kept the same (*i.e.*, 200). In the original simulation, the number of learning trials was 1000. This number is determined based on the possible number of [?] alleles in a genotype. Originally, half of a genotype is occupied by [?] alleles (*i.e.*, 10 [?] alleles per genotype). Therefore, there are 2^{10} (=1024) possible combinations of [0] and [1] in such a genotype. The number of learning trials in this replication is drawn from this number. In this replicated simulation, the number is set to 64 so that it corresponds to the proportion of the initial configuration. Strictly speaking, the proportion of the possible combination of [0] and [1] alleles in a genotype to the number of trials is larger than the original one. However, as the results show, this does not add significant differences. The fitness function is given as follows: $FITNESS = \frac{1+11n}{64}$ where n is the number of learning trials remaining after the learning process is finished. In the selection process, the "roulette wheel" selection mechanism is used[1]. Mutation is also added; the probability is 0.001 per allele. One point crossover is obligatory for every reproductive process. Otherwise stated, these configurations (*i.e.*, the selection mechanism, the mutation rate, and the crossover configuration) are applied to all other simulations give in this chapter.

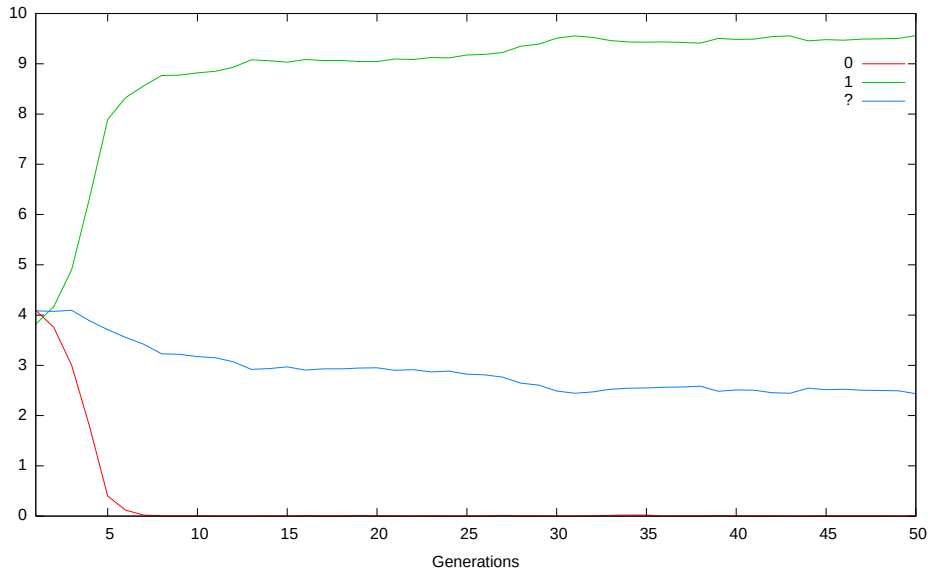

Figure 9.1: Replication of Hinton & Nowlan –First 50 Generations

The result of the first 50 generations is shown in Figure 9.1 (p. 164); the number of [0] alleles disappears almost immediately after the simulation is commenced. The

[1]Each individual is assigned a sector of a roulette wheel whose size is proportional to her fitness value. By choosing a random position on the wheel (*i.e.*, spin the wheel), selection takes place.

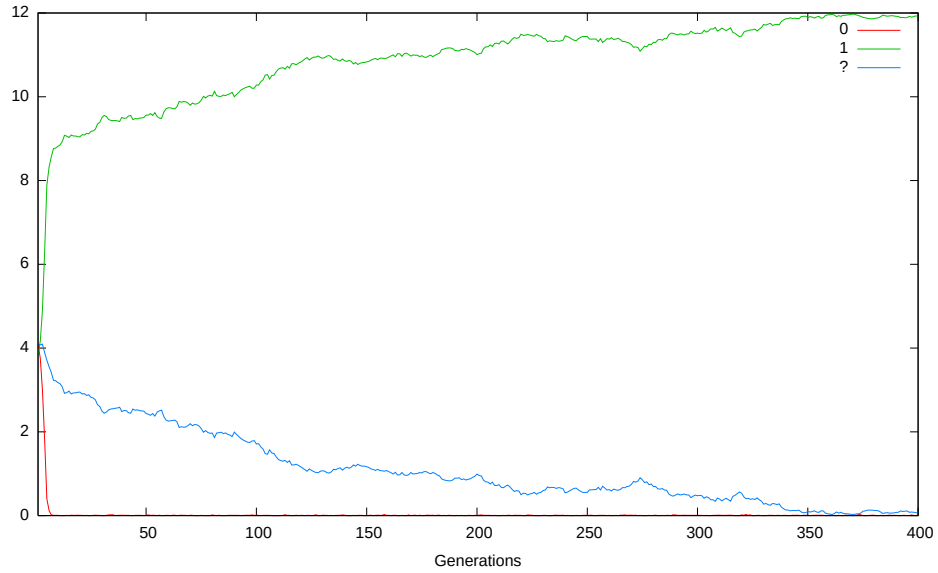

Figure 9.2: Replication of Hinton & Nowlan –400 Generations

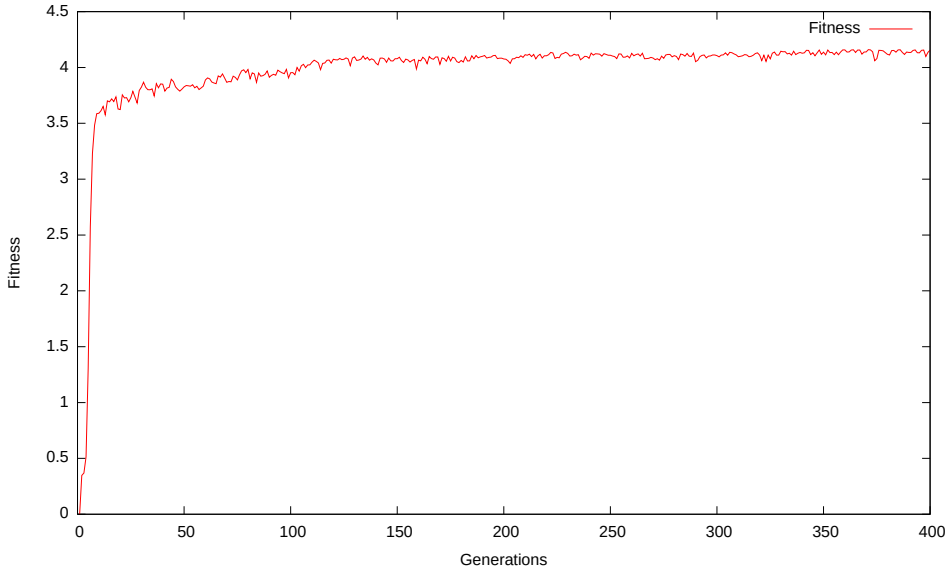

Figure 9.3: Replication of Hinton & Nowlan –Fitness

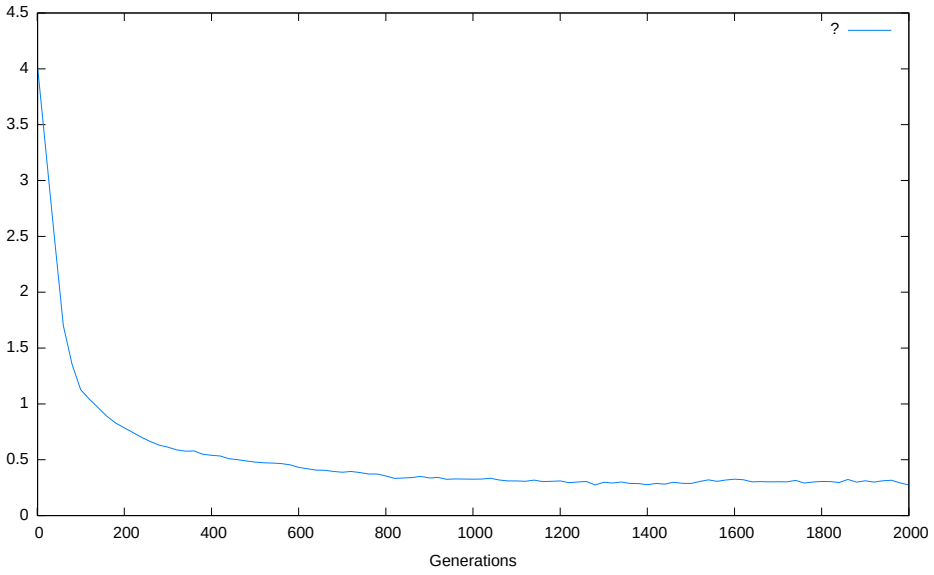

Figure 9.4: Replication of Hinton & Nowlan –2000 Generations

number of [?] alleles keeps decreasing slowly. However, as in Figure 9.2 (p. 165) where the first 400 generations of the same run are shown, the plastic allele also disappears eventually. Fitness, as expected, climbs up immediately (Figure 9.3, p. 165). Basically, the result is along the lines of Hinton & Nowlan's original result. However, the remaining number of plastic alleles is quite different; while in the original simulation, the decrescent curve is very shallow, in this simulation, the allele completely vanishes from the genepool. The reason for this seems to rest in the selective pressure, as Hinton & Nowlan originally suggested. If the mechanism is replaced with the rank selection and low-rank agents also have chances to reproduce, typically the number of [?] alleles is high at the end of a simulation even mutations are introduced. This result contradicts Arita's findings (*see* Section 2.8.3).

Finally, Figure 9.4 (p. 166) shows the averaged result of 100 runs of the same simulation which is extended up to 2000 generations. Although the number of plastic alleles does not disappear all the time, the canalizing effect is quite robust. The standard deviation at the end of the runs (*i.e.*, the 2000^{th} generation) is roughly 0.25.

9.1.2 Kirby & Hurford

As described in Section 3.2.1, Turkel (2002) conducted a simulation whose modeling architecture is based on Hinton & Nowlan (1987). In the model, in lieu of setting a fixed objective function, the function is dynamically formed; agents compare their phenotypes which are expressed from their corresponding genotypes. If a locus has a plastic allele, it randomly expresses either [0] or [1] value onto the

corresponding locus on the phenotype. If the two are exactly the same (*i.e.*, zero Hamming distance), an adaptive value is assigned. This situation can be considered to be communication establishment. If the value of the Hamming distance is more than one, they reshuffle their phenotypic values expressed by the plastic alleles and compare them once again. This process is considered to be learning. The fitness value assigned to an agent is proportional to the number of learning trials she spends before establishing communication.

Therefore, while the model keeps the basic architectural design of Hinton & Nowlan's model, the dynamic aspect of the objective function and the way it is formed make the model essentially different from the original. This dynamic formation of the function itself can be considered to be niche construction. Since this niche construction mechanism is introduced because of implementing the communicative aspect in the model, this niche construction aspect is easily conceived as a byproduct of communication, like Waddington (1975) and Dor & Jablonka (2000, 2001) consider.

Moreover, although what genotype is the best in a given population is not predetermined, because the fitness function is the same across different runs, overall fitness of different populations with different 'optimal' genotype(s) is almost identical. Thus this can be considered as a case of network externality (more specifically, consistency-over-contingencies).

Turkel considers that the genetic components of a genepool which consists of $\boxed{0}$, $\boxed{1}$, and $\boxed{?}$ alleles can be considered within the framework of the P&P model. That is, each allele is a representation of some part of our linguistic knowledge, and a genotype as a whole represents the LAD. For example, the fixed alleles, namely $\boxed{0}$ and $\boxed{1}$, correspond to principles, as these cannot be reset or overturned; principles in the P&P theory are also thought of as unmodifiable linguistic knowledge which ubiquitously appears in the world's languages. On the other hand, the plastic $\boxed{?}$ alleles are thought of parameters, because actual phenotypic values of the alleles are postnatally determined; parameters in the theory allow us to acquire different languages based upon our postnatal linguistic experiences. Since the P&P theory also generally considers that the possible range of values in both principles and parameters would be binary, this application of the P&P framework into this simulation model has an extra motivation. However, as this representation system is highly abstract, it would be better to keep in mind that the metaphor should be kept in a very abstract sense.

Based on this simple but interesting model designed by Turkel, Kirby & Hurford (1997) make a further modification. Crucially, they add a significant aspect

of language which is missing in Turkel. That is the "*linguistic inheritance*" mechanism. In Turkel, 'learning' is conducted while the agents are attempting to establish communications. This learning mechanism does not have any aspect of vertical transmission of a language; previously 'learnt' languages are discarded at the end of the generation, and new languages are formed while the agents in the new generation try to communicate. Thus in Turkel's model, while individuals are involved in niche construction, their artifacts (*i.e.*, languages) do not have a form of extragenetic inheritance. In reality, what a language learner acquires comes from previous generations and usually not from the same generation (who are not yet to speak). In this regard, Turkel's simulation has a setback in terms of implementing a notion of language evolution. By introducing a vertical linguistic transmission model, Kirby & Hurford succeed in demonstrating that this mode of transmission can significantly affect the other mode of information transmission, namely genetic transmission.

While the motivation behind their work rests in a different point from our studies, the model successfully implements a sufficiently plausible model of language evolution under a consideration of the Baldwin effect with a minimum complication. We therefore adopt our study model to Kirby & Hurford.

The basic design of the simulation is summarized as follows:

1. Spatial Organization
   Agents are distributed in a two-dimensional space. This means that both communication and learning processes are bounded by this spatial organization. The space is closed, thus each end of the space is actually connected. *See* Figure 3.1 (p. 76).
2. Population
   Basically, the design is the same as Turkel. The population size is fixed to 200. Its genepool consists of three different alleles, namely $\boxed{0}$, $\boxed{1}$, and $\boxed{?}$. The size of both genotype and grammar is 12.
3. Arena of Use
   The arena of use is simply formed by copying the whole grammatical information the 200 agents have into a reservoir. This means that the final state of every agent's grammar is copied into a linguistic pool. In the next generation, input data for learning are extracted from this pool. For the first generation, the randomly created arena is given. That is, all grammars have random values.
4. The LAD and Grammar
   This is also the same as Turkel; There are 3^{12} ($\approx 5 \times 10^5$) possible states of

LAD and 2^{12} (=4096) possible grammars exist. The initial spatial distribution of alleles in a genotype is random. Also, the number of each allele in the genotype is random. Thus, typically the genotype has four [0], four [1], and four [?].

In the model, a subset of grammars have a better chance of being learnt and increasing the agents' fitness. The learning mechanism also makes the agents be prone to have grammars in the subset. More precise explanations follow.

5. Learning and its Mechanism

   The learning is one discrete process in the simulation; during this process, no other processes will intervene (*e.g.*, fitness evaluation). In this process, all agents learn from the previous generation's data (*i.e.*, triggers) for equal times. A trigger is derived from a grammar by masking all but one bit of its information. Thus one trigger holds information of a value in the unmasked bit and its position. The unmasked position is randomly determined. In the face of the given trigger, an agent takes either of the following two different learning tactics.

   Tactic 1 is:

   (a) Compare the trigger with its current grammar.
   (b) If the grammar possesses the same value as the trigger has, then the grammar remains intact.
   (c) If there is a discrepancy, flip any one of its parameters (one [?] in the LAD), then compare it with the trigger again.
   (d) If the flipping reduces the discrepancy, adopt the new grammar.
   (e) If it fails, then keep the current grammar (the one before the flipping).

   Tactic 2 is:

   (a) Compare the trigger with its current grammar.
   (b) If those two are coherent, flip any one of its parameters.
   (c) If the new grammar is still consistent with the trigger AND the number of [1] in the first four bits of the grammar increases, then adopt the new grammar.
   (d) If there is a discrepancy, flip any one of its parameters, then compare it with the trigger again.
   (e) If the new grammar can cope with the trigger, adopt the new grammar. If not, stick to the current grammar.

   Tactic 2 models the idea that *parsability* of a grammar is counted. In each learning trial, with a probability of 0.9 the first tactic will be chosen. Thus

Tactic 2 is taken up with only a 10% chance. The basic mechanism of this learning mechanism comes from TLA with some modification (mTLA). Importantly, a learning agent can receive triggers only from its neighboring adults. The triggers come from:

(a) The previous agent who was on the same position as the current learning agent.

(b) The previous agent who was on the one immediately to the right hand side of the current learning agent.

(c) The previous agent who was on the one immediately to the left hand side of the current learning agent.

The total number of triggers for one individual is fixed to 200, while the number of triggers coming from a specific adult is not fixed. Therefore, it is theoretically possible (but very unlikely) that a learner gets all triggers from a single adult.

6. The Fitness Function

After the learning period (considered to be the critical period), each agent's fitness is evaluated by communicability of each agent's grammar. The basic mechanism of this is also similar to Turkel, but there are some significant modifications.

Each agent compares one bit of its grammar with either its left or right side agent. This procedure can be considered to be utterance and comprehension. If an utterance is comprehended by a hearer (*i.e.*, two agents have the same value at the given position in their grammars), fitness is calculated. The function is divided into two types of sub-function; the first type of the function is: with a 90% chance, fitness is increased by 1 for both the speaker and the hearer when the values of the compared bit are the same in both speaker and hearer –the utterance is accepted. With a 10 % chance, they count the number of $\boxed{1}$ alleles in the first four bits of their grammar (parsability is counted). The number directly reflects the probability of fitness increase (*i.e.*, if the agent has three $\boxed{1}$ alleles, its fitness value increases with a 75% chance).

7. The Reproduction Process

Both mutations and recombinations are introduced with the same manner as the replication of Hinton & Nowlan (*i.e.*, the probability of mutation is 0.001, and recombination is obligatory one-point crossover). The roulette wheel selection is also used. Regarding breeding, the spatial distribution of agents is ignored. This means the following two things. First, a selected

agent can freely breed with the other selected agent from any position on the one dimensional space. Secondly, their children are distributed in the space also randomly. In other words, the parents' positions are nothing to do with their children's new positions. Therefore, the mode of genetic inheritance is not affected by the spatial distribution, while that of cultural evolution is. Finally, all previous agents are wiped out in the face of the completion of the breeding process.

Results

Following Hinton & Nowlan, in this section, a replicated simulation of Kirby & Hurford is demonstrated. This time, no parsability bias is included; no bias is installed both in the learning process and the communication process. For one agent, there are 100 chances to be a speaker and another 100 chances to be a hearer. So, the highest fitness value one can have is 200. The lowest is one. The frequency of [0], [1], and [?] alleles in the initial genepool is equal (*i.e.*, roughly four each in a genotype), while in the original, it is fully occupied by [?] alleles. The selection mechanism is the roulette wheel selection mechanism (in the original, the rank selection is used; the top 90% of the population have an equal chance of reproducing). The result is shown in the figures (from Figure 9.5 to Figure 9.8).

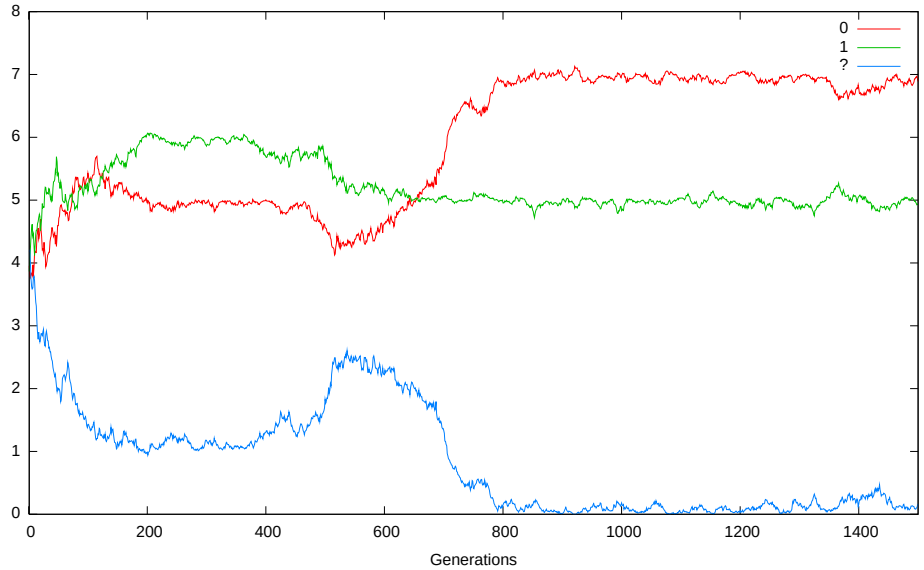

Figure 9.5: Replication of Kirby & Hurford –1500 Generations

The result obtained from the replications is, as expected, almost identical to Kirby & Hurford's original simulation. Figure 9.5 (p. 171) shows the evolutionary trajectory of [0], [1], and [?] alleles in a typical run. The number of [?] alleles quickly goes down to 0. As seen in Figure 9.6 (p. 172), the average fitness increases

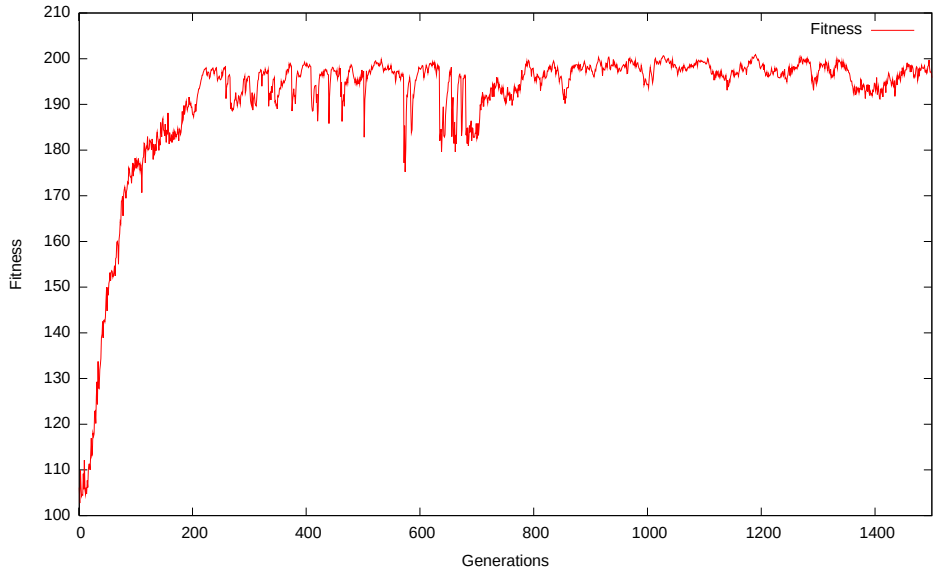

Figure 9.6: Replication of Kirby & Hurford –Fitness

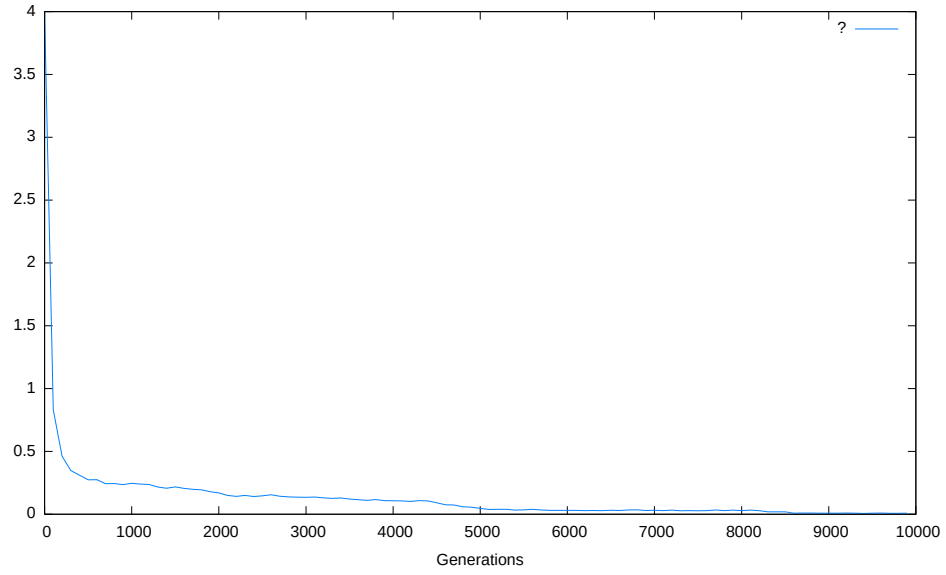

Figure 9.7: Replication of Kirby & Hurford –Averaged Result of 100 runs

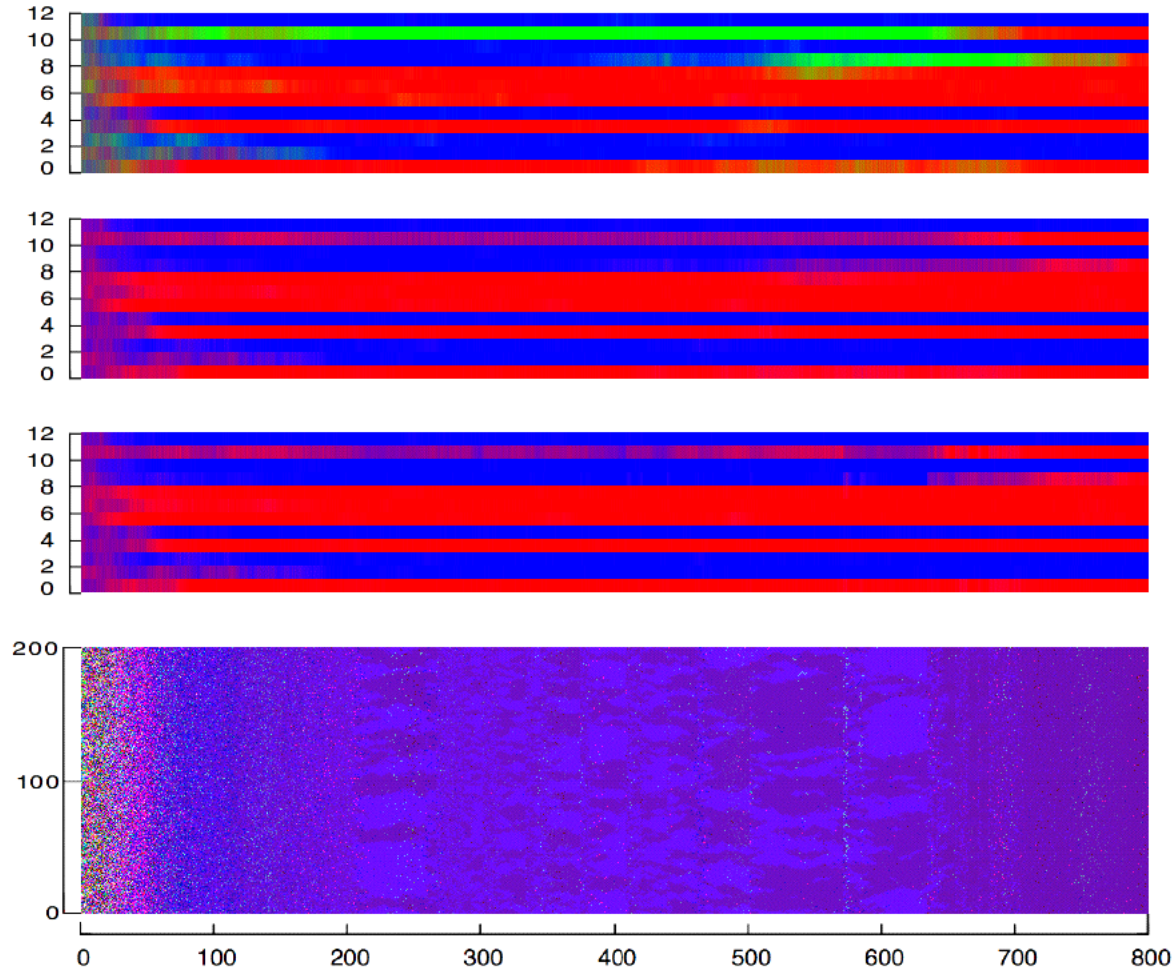

Figure 9.8: Replication of Kirby & Hurford –A spatiotemporal graph

along the line of this loss of [?] alleles. This implies that the whole population converges to almost a single genotype. This can be also seen in Figure 9.8 (p. 173). In the figure, the first stripe shows the graphical representation of an averaged LAD (*i.e.*, genotype). The second is pre-learnt language (*i.e.*, before learning with *m*TLA takes place), and the third is learnt language. The stripe is divided into 12 threads, and each thread designates a corresponding locus in a grammar. [0] grammatical information (*i.e.*, [0] allele in the grammar) is colored red and [1] is in blue. Green is the plastic allele. All stripes are, again, averaged results of all agents in a population. Therefore, if the dominant color is either red or blue, this means that the majority of the population has the same grammatical information on the specific locus. On the other hand, a purple region implies that the population is divided at that region. The first locus is allocated on the bottom of the stripe and the last locus on the top. Time runs horizontally from left to right. The figure shows from the initial generation up to 800. Note that, around the 600^{th} generation, in

the LAD, the region covered by green increases slightly and quickly fades around the 770^{th}. This is also found in Figure 9.5 (p. 171).

The last, large picture shows the spatiotemporal distribution of the grammars in the population. Each grammar is assigned an idiosyncratic color. Grammars obtained from Agent 1 to Agent 200 are allocated vertically from the bottom to the top. Initially, the agents have individually different grammars. However as time goes by, a patterned structure appears. This means that through the learning process, languages are inherited by learners who are neighbors of the adults. However, because of mutations, some sharp surges take place during the evolution. This reshuffles the distribution of languages. This corresponds well to the surges on the fitness value (Figure 9.6, p. 172). Finally around the 770^{th} generation, where the plastic alleles disappear, the color goes to monotonic; a single grammar almost dominates the population (apart from sporadic appearances of mutated grammars).

In summary, the simulation successfully demonstrates that the Baldwin effect takes place in language evolution under a comparatively simple assumption even with a linguistic inheritance mechanism.

Note that in Kirby & Hurford (and also in Turkel), the genetic basis of the plasticity is also in an allelic relationship to the fixed linguistic knowledge. Moreover, they are positively correlated; as the number of plastic alleles decreases, the degree of plasticity diminishes at the phenotypic level. However, it is equally important to note that, in these models, the mechanism of the Baldwin effect is fundamentally equivalent to BNC; through establishing an internal norm (*i.e.*, grammars with which agents can communicate), agents can increase their fitness. This norm creation is a niche construction process. Therefore, importantly, while the model is based on the G-P correlation model as an adoption from Hinton & Nowlan, it is also a model of BNC.

9.2 Implementation of Epistatic G-P Decorrelation

Through the replications of Hinton & Nowlan and Kirby & Hurford, it should be clear now that both the canalizing effect and the expediting effect clearly take place in these types of GA models. In this section, we examine how different types G-P decorrelations affect the Baldwin effect by modifying the above models.

9.2.1 NK-Landscape Model

The first type of G-P decorrelation is related to epistatic relationships. By introducing intragenomic epistatic interactions, the linear G-P relationship can be disturbed; as a given allele's expression is determined by other alleles in different

loci, a selective process on the phenotypic level would not be linearly reflected on the genotypic level.

In the following modified simulations, this 'context-dependent' model of the gene expression mechanism is implemented by Kauffman's (1989) *NK*-Landscape model. In the *NK*-Landscape model, unlike ordinal GA models where one gene expresses one trait of a phenotype, a *set* of genes 'non-linearly' determines a trait in a phenotype. In other words, one trait may be decided by two or more distinctive genes. How many genes are required to express one trait is controlled by the value of K. The values are always between 0 and N-1 where N designates the number of the genes[2].

Dependency of genes can be either "*contiguous*" or "*non-contiguous*". In the case of contiguous dependency, a gene forms a concatenation with other adjacent genes. Note that in the contiguous dependency case in a computer simulation, both ends of a genotype can be considered as neighbors to each other so that K-dependency of phenotypes is available in all loci for a practical reason. In the non-contiguous dependency case, on the other hand, the group of genes is randomly dispersed. In this thesis, only contiguous cases are considered.

It is clear that as the value of K increases, dependency between different genes increases. In terms of evolutionary search, the increase of the value of K means increase of the degree of epistasis; the fitness landscape becomes progressively rugged. In a rugged landscape, evolutionary search tends to get trapped in local optima. This is a case in which the correlation between genotypes and phenotypes becomes low. From the perspective of a focal gene, its allelic value is not directly reflected on the phenotypic level since is it has to be non-linearly determined with K-1 other genes in a given genotype. Under such a condition, substitution of alleles by genetic operations becomes less and less correlated to possible modifications of phenotypic values.

A given combination of genes consisting of K different alleles expresses a predetermined phenotypic value onto a corresponding position on a grammar. This does not change throughout the run. More concretely, in our model, a set of alleles, say $\boxed{001}$, will express $\boxed{?}$ onto a given position in the corresponding grammar. This expression mechanism is done by having a randomly generated expression table at the beginning of a simulation run, and will not change across the same run[3]. The

[2]Theoretically, it is possible to model that the maximum value of K is equal to N. However, as this does not add any meaningful insight for a GA model, here we consider up to the case of $N>K$.

[3]It is possible to design a simulation in which K is dynamically changed in a run. However, as it is beyond the scope of this thesis, we do not discuss here.

Locus \ Alleles	**0**00	**0**01	**0**10	**1**00	**0**11	**1**01	**1**10	**1**11
L1	0	?	1	1	?	?	0	?
L2	?	?	?	0	1	1	0	?
L3	1	?	?	0	1	0	?	?
L4	?	0	1	0	?	?	?	1
L5	?	1	?	1	0	0	?	?
L6	?	?	?	0	0	1	1	?
L7	?	1	0	?	?	1	0	?
L8	0	?	1	0	1	?	?	?
L9	?	?	0	0	?	1	1	?
L10	1	?	0	?	0	1	?	?
L11	?	1	0	?	?	0	?	1
L12	?	1	1	0	?	?	0	?

Table 9.1: A Look-Up Table –K=3

table can be a considered as a look-up table of a gene expression whose size corresponds to N times 2^K since each allele is affected by 2^K possible combinations of other genes.

For example, suppose we conduct a simulation in which organisms have 12 genes in their genotype. Thus N=12. Subsequently, possible values of K ranges from one to 11 (*i.e.*, $0<K<N$). Then, a look-up table is generated. This table specifies a phenotypic value from a certain set of alleles. An example is shown in Table 9.1 (p. 176):

The number of rows corresponds to N (*i.e.*, 12). The number of columns corresponds to the number of possible combinations of genes. In the table, the value of K=3; there are 2^3 (=8) possible combinations. When K=11, the number of combination will be 2^{11} (=2048). Each cell is filled with a fixed or plastic allele (*i.e.*, $\boxed{0}$, $\boxed{1}$, or $\boxed{?}$ allele). The corresponding value is mapped onto a designated locus of a grammar. Note that the representation of plasticity is implemented in the level of the phenotype but not directly in that of the genotype. This is the reason that the value of the gene is binary (*i.e.*, two types of alleles –$\boxed{0}$ and $\boxed{1}$).

Finally, this type of *NK*-model implements both epistasis and pleiotropy at the same time, as a gene expresses a phenotypic trait with the context of other genes (*i.e.*, intragenomic epistasis), and also contributes to multiple numbers of phenotypic traits (*i.e.*, pleiotropy).

9.2.2 The Model

To make the simulations comparable to the replicated simulations, in the following simulations, most parts of the original architectures are preserved. The frequency

of [0], [1], and [?] alleles is set to be equal. This means that in a look-up table, statistically equal numbers of [0], [1], and [?] alleles appear on a row. Therefore, in a look-up table of K=N-1, there are roughly 683 ($\approx \frac{2048}{3}$) alleles for each type of allele appearing on a single row of a look-up table. Since [0] and [1] alleles are randomly distributed on an initial genepool, the distribution of different phenotypes also follows a Gaussian distribution.

At the beginning of a generation, all agents express their LADs (*i.e.*, principles/parameters) based on their genotype according to a given look-up table. The number of phenotypic traits in one agent is the same as the number of loci of its genotype. To express 12 principles/parameters, a translator reads a genotype from locus 1. In the case of K=3, for example, the translator first reads the alleles in loci 1, 2 and 3, and produces one principle/parameter at the first position in the corresponding phenotype. This process proceeds in an iterative fashion; the translator reads locus 2, 3, and 4, and puts a principle/parameter on locus 2 of the phenotype. If the translator reaches locus 11, the third gene is beyond locus 12. In this case, the translator refers to the gene of locus 1. The first locus of any three succeeding loci corresponds to the locus of its phenotypic expression. The locus is called a "head". When the translator reads from locus 3, for example, it refers to the third column of a look-up table. The corresponding gene of a principle/parameter is the "head" of a set of genes.

9.2.3 The Simulations

In this section, we investigate the results of G-P decorrelation implemented by the *NK*-model with the simulations given above. With the two original simulations (*i.e.*, Hinton & Nowlan, and Kirby & Hurford), three configurations –K=3, 6, and 11, are tested.

Hinton & Nowlan

First, three results of Hinton & Nowlan are shown (from Figure 9.9 to Figure 9.11, pp. 178-179). In all cases, the Baldwin effect is well suppressed; the canalizing process is effectively blocked. Epistatic decorrelation seems to prevent the population from reducing the number of the plastic alleles.

However, the result is little more complicated. Even in the case of K=11, the number of [0] allele falls to zero, while the plastic allele has not been reduced so effectively. The possible reason is that [0] allele is completely deleterious while the plastic alleles are potentially adaptive. Therefore, during the initial stage of evolution, those who can get rid of the deleterious allele become (potentially) adaptive. However, as such a reduction process goes on, the genetic diversity is lost so

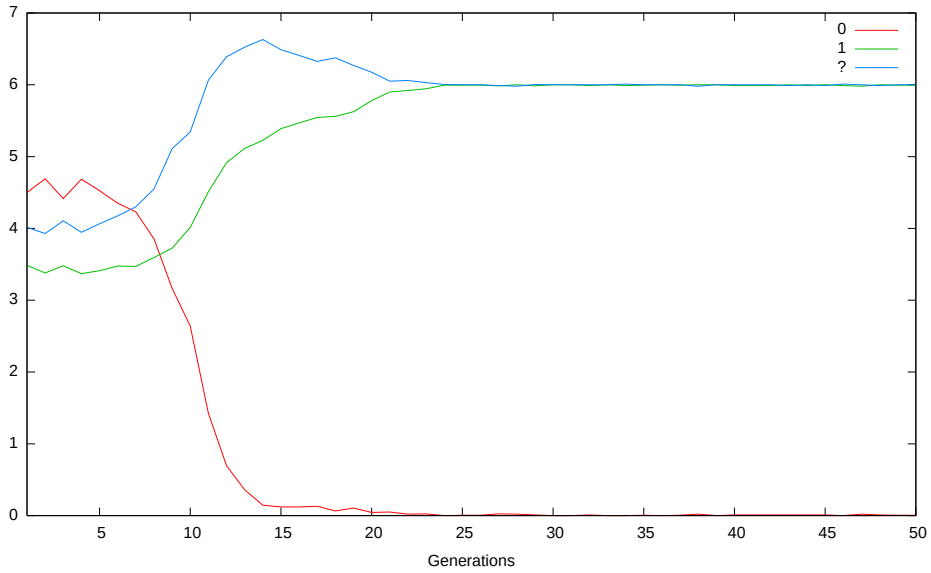

Figure 9.9: Hinton & Nowlan: $K = 3$

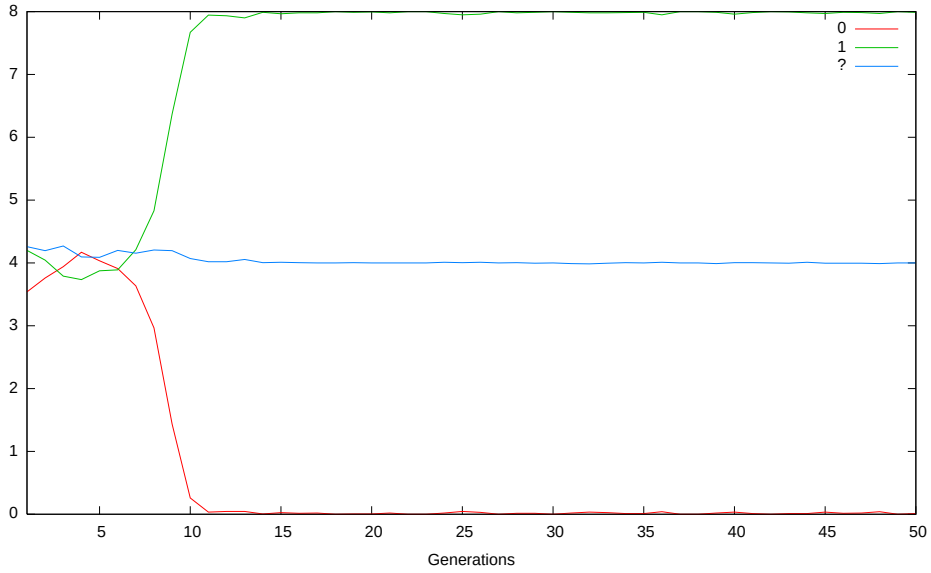

Figure 9.10: Hinton & Nowlan: $K = 6$

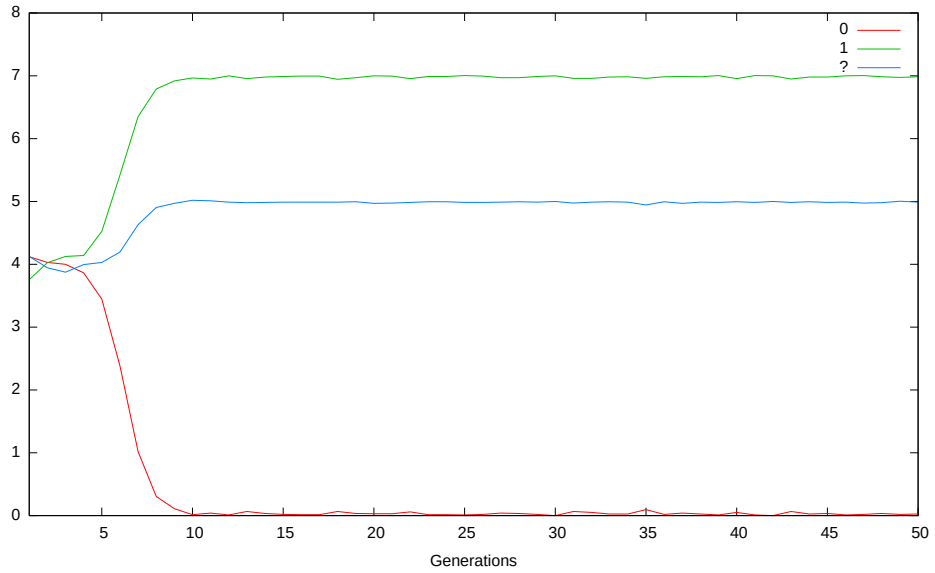

Figure 9.11: Hinton & Nowlan: $K = 11$

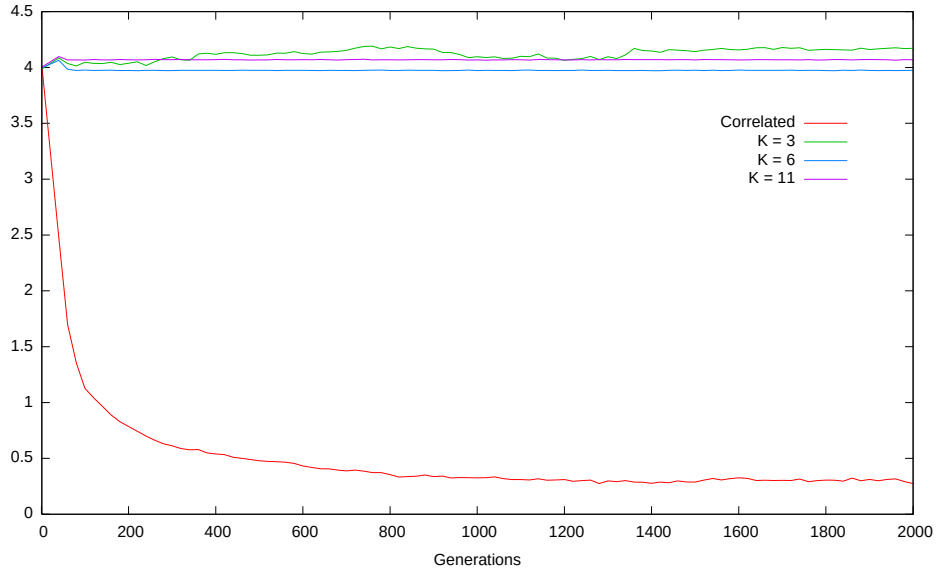

Figure 9.12: Hinton & Nowlan: The Averaged Results of 100 runs

that reshuffling by crossovers cannot produce enough diversity in addition to the epistatic condition.

Another puzzling fact is that across the different values of K, the degree of the Baldwin effect is almost the same. For example, when the remaining plastic alleles are measured at the end of 300 generations with 100 runs, the averaged results are 3.5, 2.9, 4.2, under the configurations of K=3, 6, 11, respectively (Figure 9.12, p. 179). A hint may be found in the standard deviations. The deviations are 1.58, 1.10, and 0.63, respectively; as the value of K increases, the deviation gets smaller. With small values of K, genotypes may be 'genetically' canalized[4], while in high K values, the entire evolvability of the population is reduced. In other words, with high K values, there is only small room for evolutionary search to move. This would be reflected in the small standard deviations. A more detailed explanation is given later.

Kirby & Hurford

The first set of figures (Figure 9.13 –K=3, Figure 9.14 –K=6, and Figure 9.15 –K=11, pp. 181-182) show the results of Kirby & Hurford with the epistatic G-P decorrelation. As reflected in the graphs, the Baldwin effect is suppressed in all of these figures; in all figures, the average number of plastic alleles in a genotype remains the same as the initial number. In other words, no canalization process takes place.

Notably, while in K=3, some evolutionary dynamics are observable, in K=11, the whole genepool is immediately occupied by a single (or very small numbers of) genotypes so that no particular dynamics takes place. As the configuration of K=6 demonstrates, some minor dynamics would take place between these two extreme values (*i.e.*, K=3 and K=6). This implies that under low values of K, genetic operations can produce some diversity in a genepool, while in higher values, this is effectively blocked.

When the averaged results of 100 runs of each condition are examined (Figure 9.16, p. 182), it becomes clear that the number of plastic alleles slightly decreases in lower values of K, while when K=11, almost no (environmental) canalization takes place.

9.2.4 No Mutation, No Recombination

In the above section, the results of the simulations demonstrate that G-P decorrelation by epistasis and pleiotropy indeed blocks the Baldwin effect. This is what

[4]Note that so far, we have used the term "canalized" as equivalent to "environmentally canalized". However, in this particular case, what is canalized is not $\mathbf{G} \times \mathbf{E}$ norms of reaction, but intragenomic polygenetic norms of reaction; genetic canalization (*see* Chapter 2.)

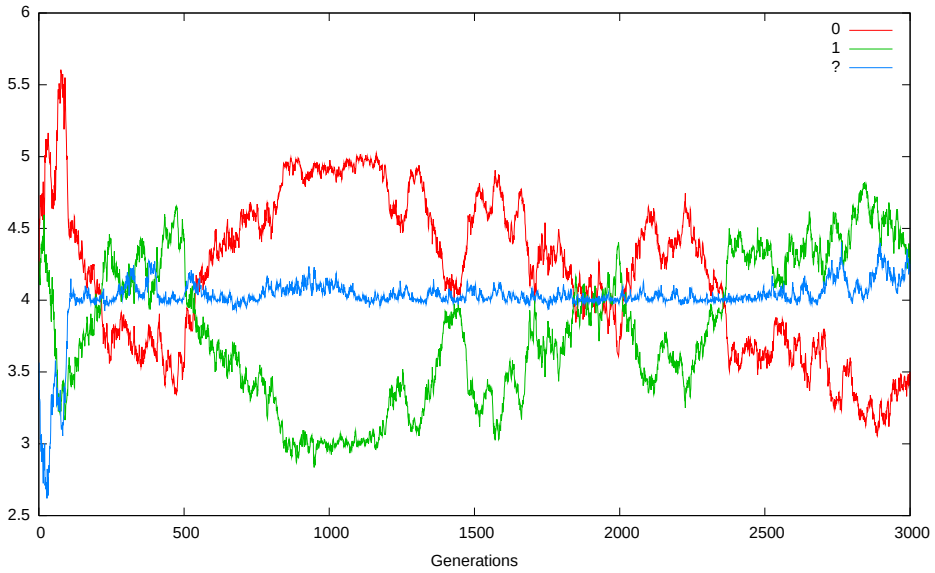

Figure 9.13: Kirby & Hurford: $K = 3$

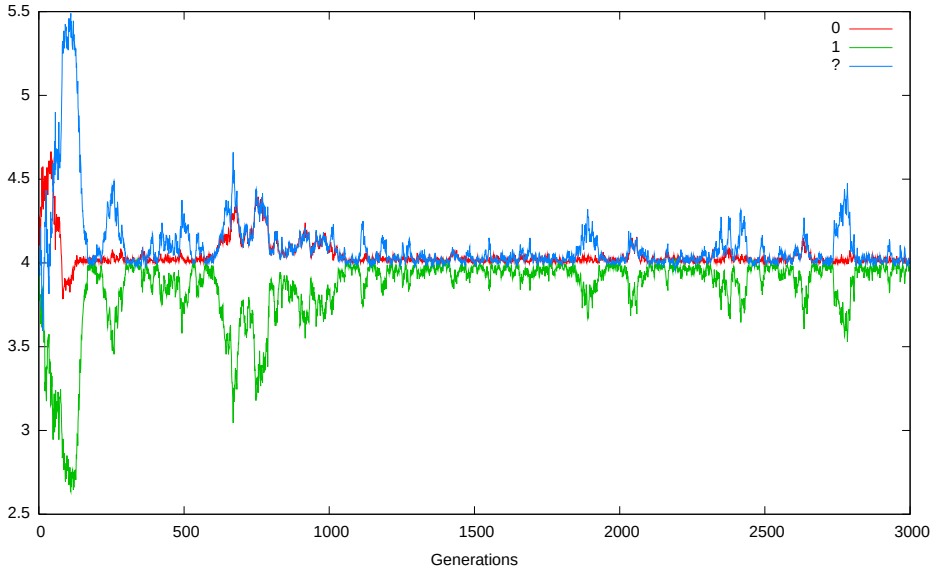

Figure 9.14: Kirby & Hurford: $K = 6$

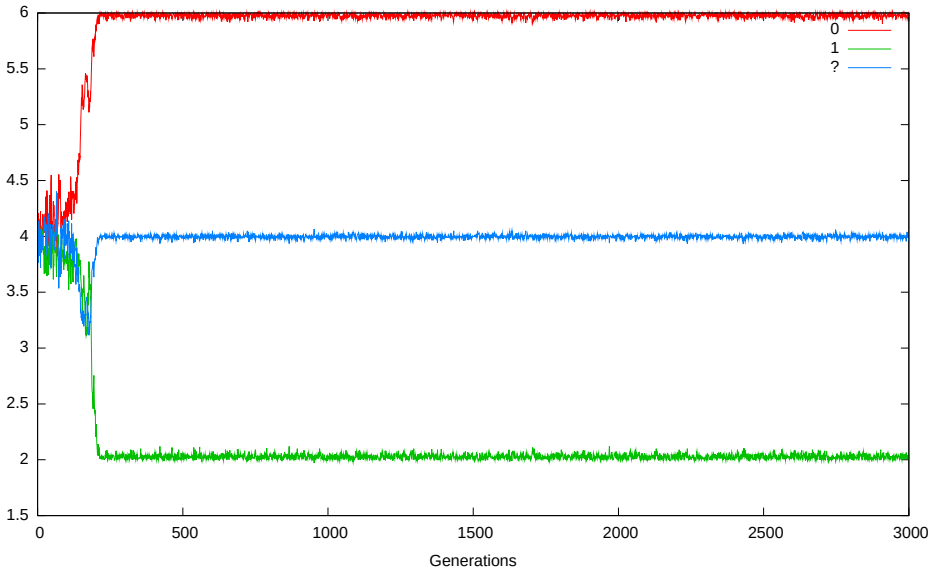

Figure 9.15: Kirby & Hurford: $K = 11$

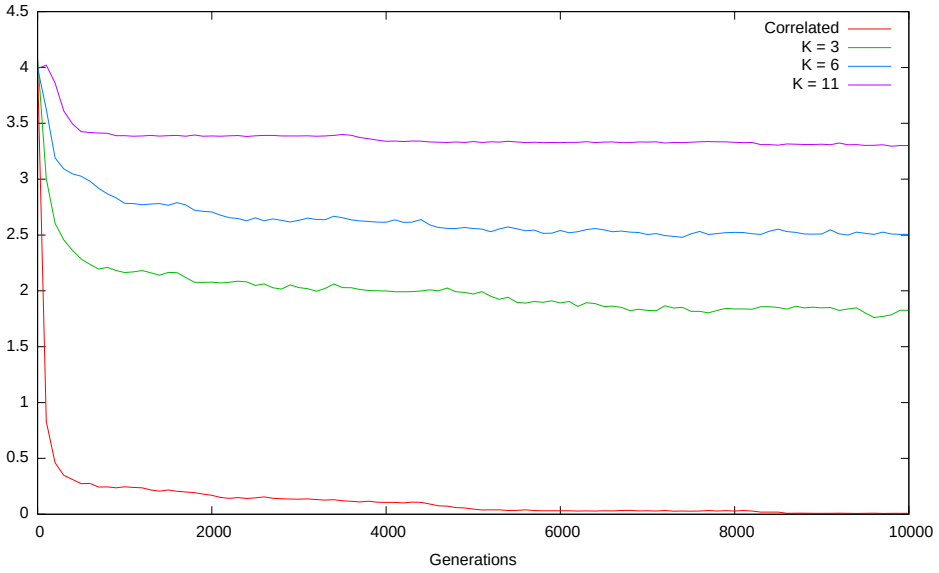

Figure 9.16: Kirby & Hurford: The Averaged Results of 100 runs

Mayley (1996*b*) and Yamauchi (1999, 2001) present. However, this is a somewhat perplexing conclusion; as with the simulations of language evolution, where BNC is supposed to be involved, the Baldwin effect is blocked with a considerable degree.

Two factors would be involved here. The first one is the size of context. Recall the discussion given in Chapter 2. In Section 2.4.1, we see (intragenomic) epistasis is created through genetic canalization. The blood type was given as an example of dominant and recessive epistasis there. Suppose there are just two types of alleles in a genotype (say, a^1 and a^2, respectively). If they are in the simple dominant-recessive relationship, regardless of the size of K, the number of context (*i.e.*, the number of phenotypes) is just two; either the dominant allele(s) 'dominates' or not.

Suppose then, the size of K is two (while the size of N can be arbitrarily long), and also they are not in the dominant-recessive relationship. If order of allelic allocation is not important, there are 2^2 possible combinations of a set of two genes creating a phenotypic trait while the number of different phenotypic values is three; $< a^1a^1 >$, $< a^1a^2 >$ (or $< a^2a^1 >$), and $< a^2a^2 >$. Therefore, if the size of K is small, the actual number of available phenotypic value may be highly biased. In other words, organisms are highly genetically canalized. This would be the reason that in such small K, the standard deviation is comparatively large across different runs.

The second factor is genetic operations. For example, genetic recombinations often create the hitchhiking effect. When the size of K is comparatively small, sets of alleles forming phenotypic units can be often hitchhiked without breaking the combination. In other words, the cutting point of a crossover does not match most of the sets; most of the sets are just moving within a genotype without breaking their bond. However, as the size of K increases, almost every single set of alleles is inevitably affected by the process. When K=N-1, with the probability of $\frac{1}{2(N-1)}$, only two points –both leftmost and rightmost cutting points can incorporate a set of phenotypic unit whose head rests on either the rightmost or leftmost position, respectively. Therefore, reshuffling by recombination in high K values is extremely deleterious.

Mutation is also the same. The process can add some genetic variances onto a given genepool. Having said that, as K increases, one mutation can influence many phenotypic units simultaneously. If K=N-1 again, one mutation can affect everything but one.

Therefore, it is conceivable that even if niche construction can provide a good exaptation effect, everything is messed up by these types of genetic operations. Suppose a genepool has a sufficient variance, it would be more advantageous if the genepool was not disturbed by them. To test this assumption, the model of

Kirby & Hurford is once again modified so that no genetic operation takes place. In this model, apart from removing recombinations and mutations, everything is kept the same. Under a normal evolution, without genetic operations, the initial genetic variance would not sufficiently direct the whole population to a favorable (*i.e.*, adaptive) state. On the other hand, as a type of exaptation process, niche construction may well efficiently expose previously not-so-adaptive genotypes even without genetic operations.

In the modified simulations, as no genetic operation is implemented, reproduction is simply done by copying the genotypes of selected agents without any crossovers or mutations. Therefore, the whole population simply utilizes the initial genetic diversity. Because the number of possible LADs in the gene length of 12 is more than 5×10^5 (3^{12}), to reduce the number, in the simulations, gene length is decreased to 8 so that the possible states of LAD is less than 7000 ($3^8 = 6561$).

In the first simulation, the population size is kept the same as the original, namely 200. Therefore, the initial population can cover roughly 3% of the whole genotypic space. In the second simulation, the population size increases to 4000. This time, the initial population would cover roughly 60% of the space. In the third, the size is further increased up to 8000 which statistically covers every possible configuration of the LAD. The summarized results of the two simulations are given in Figure 9.17 (p. 185). Again, the results are the average of 100 runs. As the figure shows, the number of plastic alleles is more strongly eliminated in the larger population. The result clearly indicates that niche construction can successfully exaptate the extant genetic variation so that the population evolves to a favored state.

9.3 G-P Decorrelation by Discrepant Demands

In this section, a different type of G-P decorrelation is investigated. Recall that in the original study of Kirby & Hurford, a type of linguistic bias is implemented both in the learning and communication processes (*see* Section 9.1.2). That is, during the learning period, with 10% chance, grammars which have $\boxed{1}$ in the first four loci are preferred over the other grammars. More precisely, when an input is accepted by a current grammar, with 10% chance, the biased mTLA algorithm randomly flips one of the parameters a given agent holds. If this happens to increase the number of $\boxed{1}$ in the first four loci on its grammar, and still the input is acceptable, the algorithm keeps the updated grammar. Otherwise it resets the update point so that the previous grammar is retained. Kirby & Hurford interpret this as linguistic parsability in language acquisition. Through generations, this preference is reflected on existing languages so that such languages are 'streamlined' to the preference of

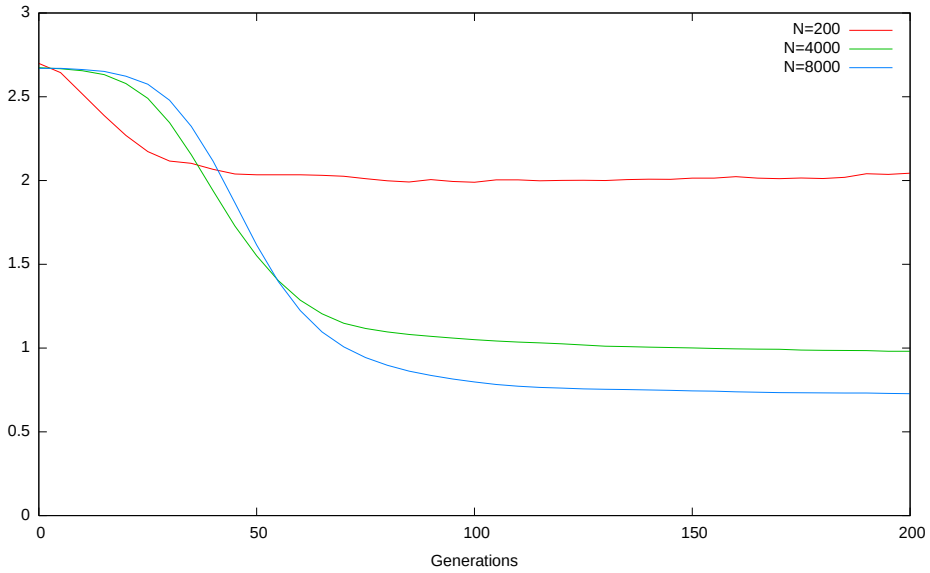

Figure 9.17: The Averaged Results of No-Mutation

language acquisition. They also set the same preference on communication; when communicability is calculated, the number of $\boxed{1}$ is checked in the same manner described above. This time, such language parsability is related to communication.

What they find is that when this parsability mechanism is implemented, it is effectively reflected on the LAD. In other words, through the canalizing effect, such a preference becomes highly ineluctable (in the simulation, it is encoded in genes). However, interestingly, when the parsability preference in language acquisition is disabled and that in communication is retained, no obvious Baldwin effect is observed. This implies that the language acquisition process has a greater influence on evolution of the LAD than the communication process can impose. In other words, natural selection which is directly influenced by communicative successes in the model is relatively powerless regarding shaping the LAD.

Suppose there is a discrepancy of parsabilities between language acquisition and language communication, it can be considered as a type of G-P decorrelation; as communications ultimately relate to genetic optimization, while the preference in language acquisition is a type of learning optimization; if these two demands are different, the Baldwin effect may be disturbed. This is somewhat similar to the situation which is considered by Best (1999); recall that he demonstrates that when social learning points in a different direction from that of individual learning, the Baldwin effect is blocked when the discrepancy becomes extreme (*see* Section 2.8.5). However, as this time the discrepancy rests along the line of learning and adaptability, it is more suitable to be considered as a case of G-P decorrelation described in Chapter 4.

To examine this type of decorrelation, two types of simulations are conducted. The first is a replication of Kirby & Hurford with the linguistic bias. Different from other replications, in this simulation the initial genepool is fully occupied by [?] alleles. This is the same as the original configuration of Kirby & Hurford. The size of a genotype is 12; larger than the original (*i.e.*, the gene length in the original simulation is 8). The population size is 200. Therefore, apart from the initial number of [?] alleles in the genepool, and the gene length, everything remains the same as the replication of Kirby & Hurford.

The number of inputs for one agent is 200, and the number of communicative attempts is 100. Again, this is the same as the original. Both linguistic and communicative biases are implemented as in the same manner; with 10% chance, check the number of [1] in the first four loci of a grammar, and according to the number, parsabilities both in acquisition and communication are determined.

In the original configuration, the configuration of the optimal parsability both in acquisition and communication is the same, is called "*positive*" as their biases are positively correlated. On the other hand, in the current configuration, while the optimal configuration of the parsability in language acquisition remains the same, that of communication is set to be opposite; the more [0] in the first four loci, the more chance the agents have to increase their fitness. This configuration is termed "*negative*".

The results of the positive configuration are shown in the following figures (Figure 9.18 and Figure 9.19, p. 187). As in Figure 9.18, [?] alleles completely disappear around generation 3000 in the positive configuration. The following three colored figures in Figure 9.19 (p. 187) are the visualization of the evolution. Similar to Figure 9.8 (p. 173), the first band represents the averaged genotype, the second and the third are the averaged pre-learnt and post-learnt grammars, respectively (to fit the whole 3000 generations, the figures are horizontally scaled). The first four positions of the genotype (shown in the four threads from the bottom of the top figure) are initially occupied by plastic genes (green). However, soon the alleles are replaced by [1] (blue). In the grammar level, even from the beginning, learning makes sure that all agents learn [1] in the first four loci. This appears in the third band. Notice that in the second band, which shows the pre-learnt grammar state, includes some purple regions in the first four threads. These regions correspond to the genic loci which are occupied by the plastic alleles. However, as in the corresponding positions in the third band, these parameterized regions are properly colored with blue ([1]). Finally, the fourth (spatiotemporal) graph shows that grammars are spatially organized. However, as the plastic alleles disappeared around generation 2700, the

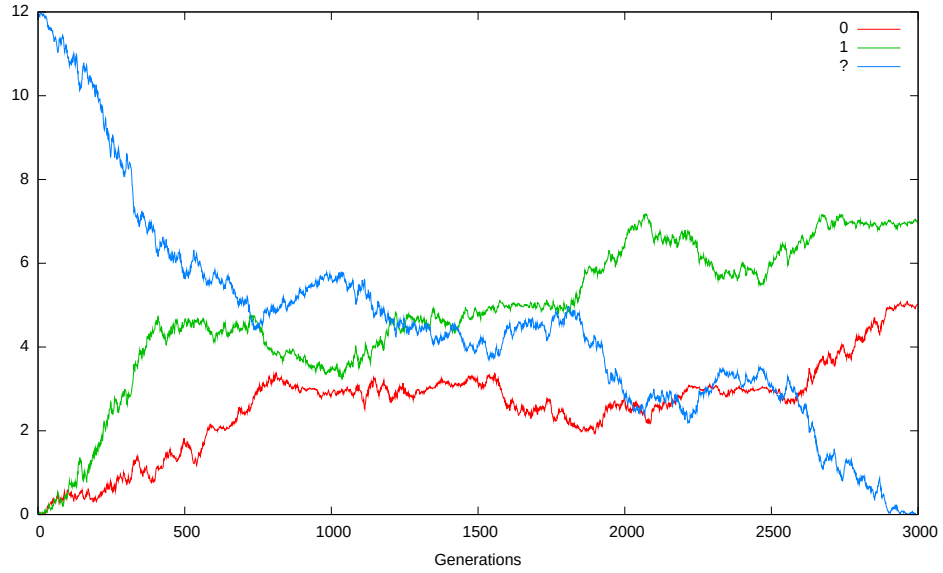

Figure 9.18: Positively Correlated Biases

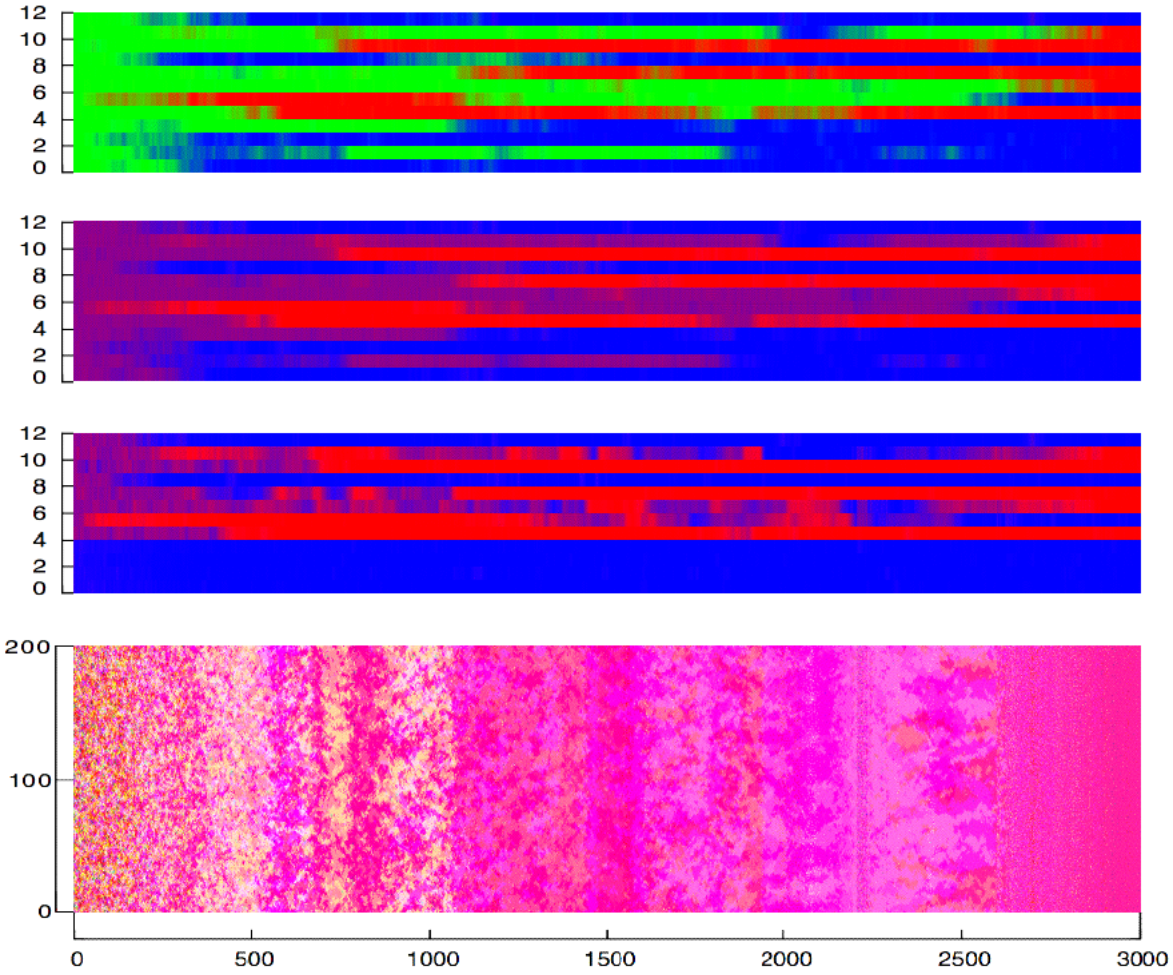

Figure 9.19: Spatiotemporal Graph of Positively Correlated Biases

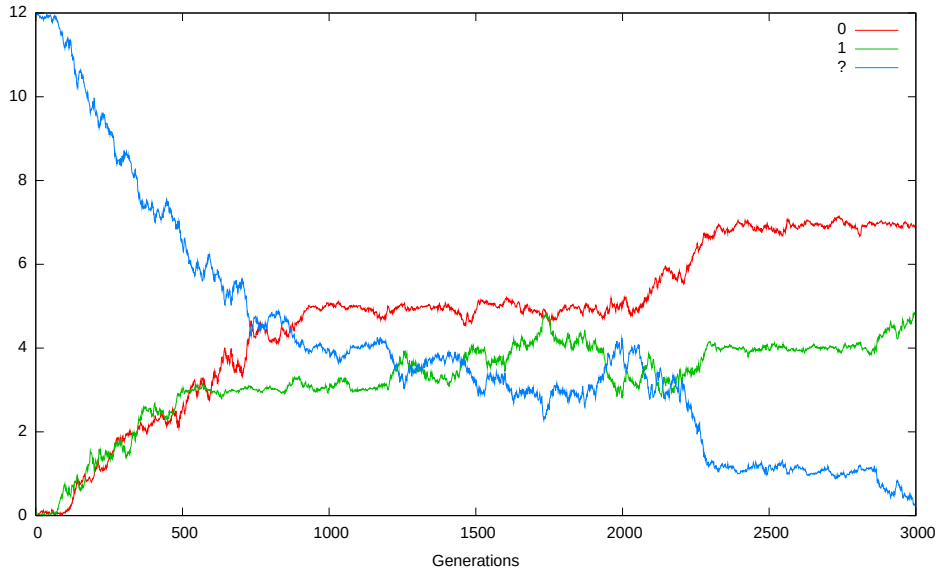

Figure 9.20: Negatively Correlated Biases

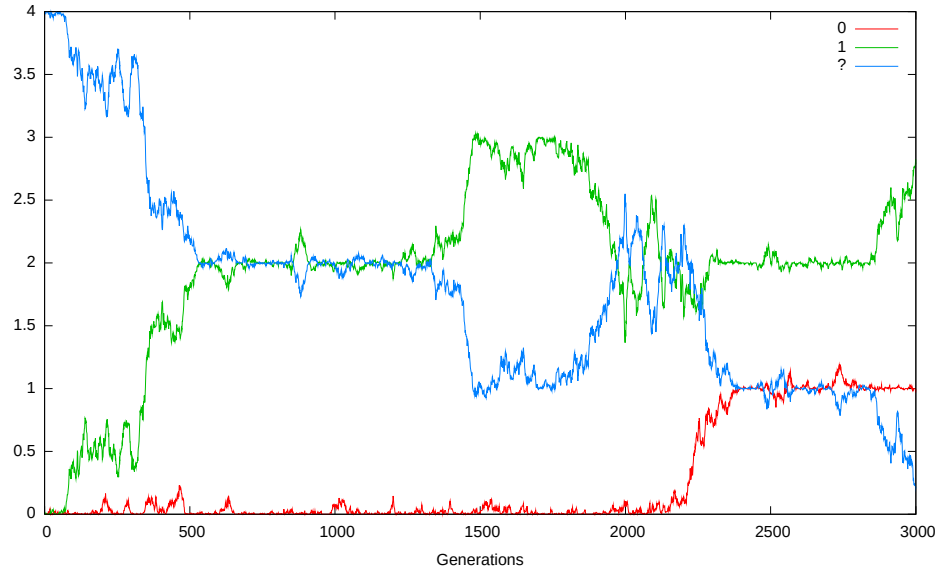

Figure 9.21: Negatively Correlated Biases in the First Four Bits

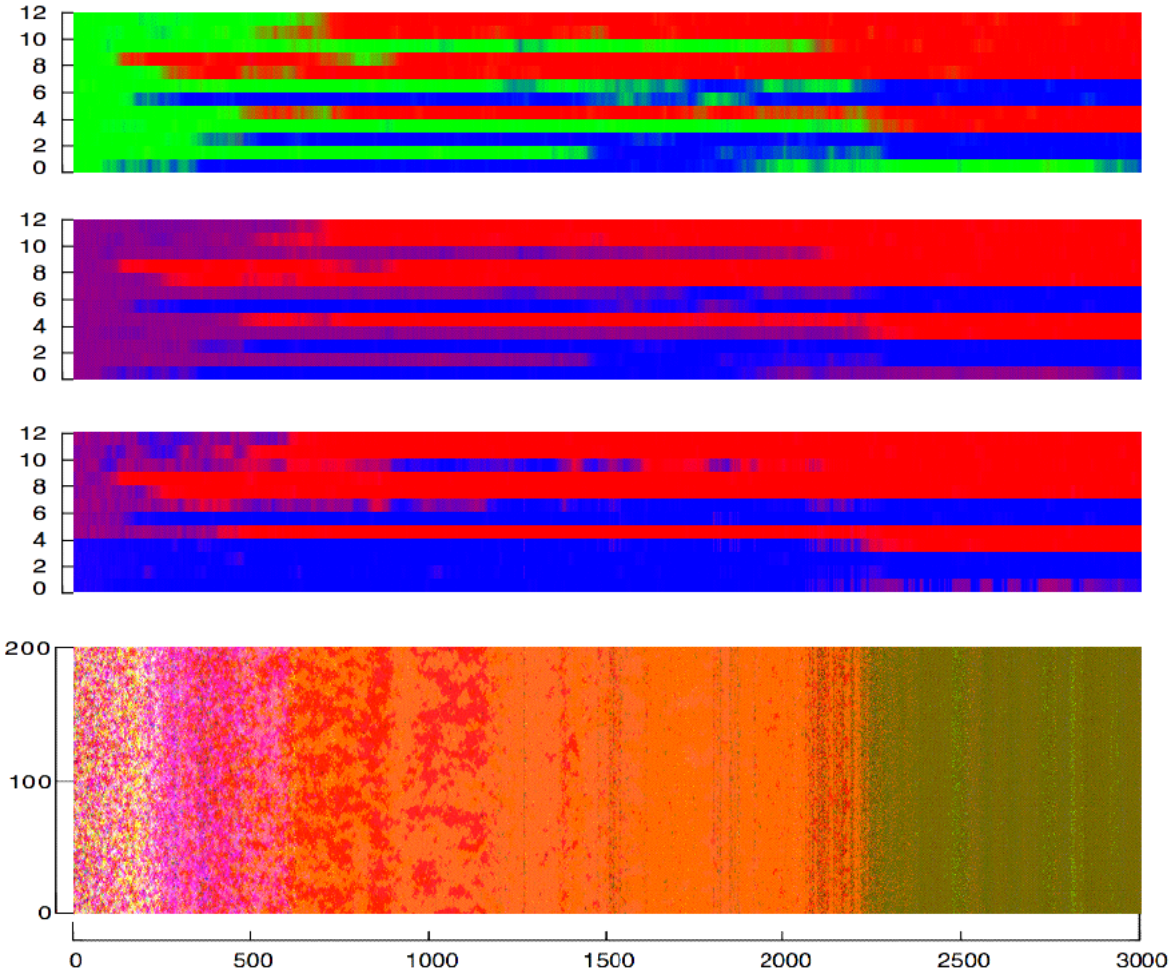

Figure 9.22: Spatiotemporal Graph of Negatively Correlated Biases

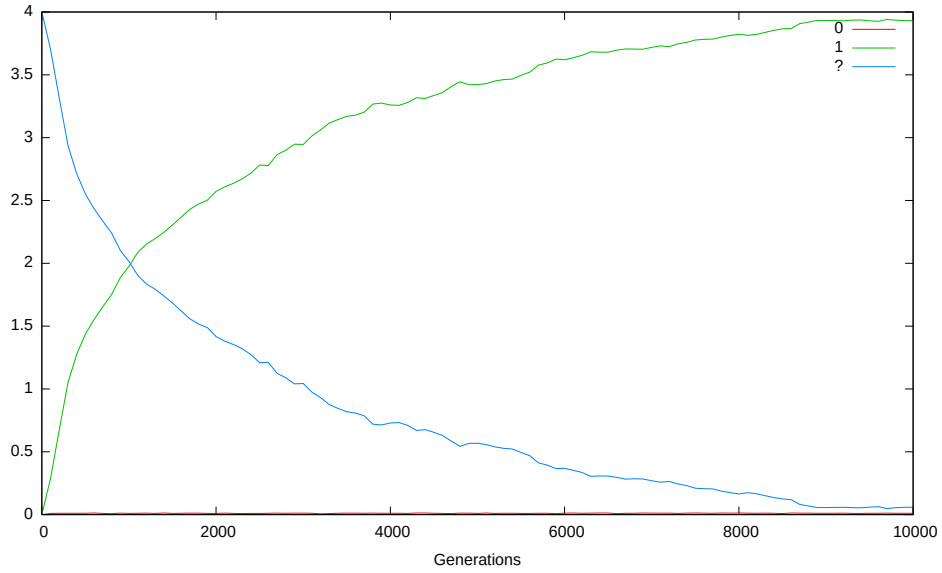

Figure 9.23: The Averaged Results of Positive Correlation

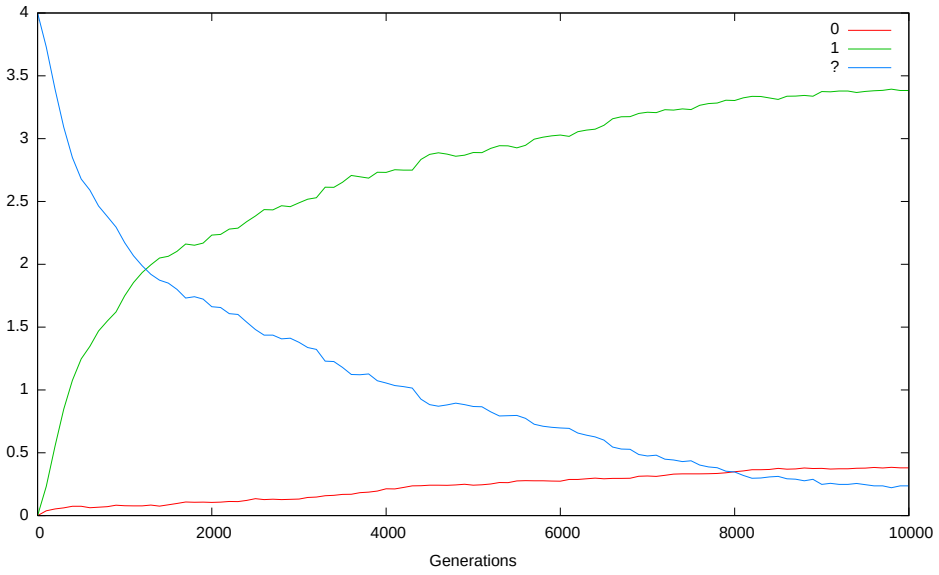

Figure 9.24: The Averaged Results of Negative Correlation

whole population is occupied by a single grammar. This is reflected in the figure; no pattern, but monotonic color appears in the corresponding region.

When the biases are configured so that they are competing with each other –G-P decorrelation, a different result emerges (from Figure 9.20 to Figure 9.22, pp. 188-189). In the second simulation, the parsability preference in language acquisition is kept to [1], while that in communication is set to [0]. The rest of the model is left intact. As in Figure 9.20, the overall evolutionary trajectory is similar to the original; parameters steadily decrease as generations pass. However, when the first four loci are focused, it depicts quite a different picture. Figure 9.21 is taken from the same run shown in Figure 9.20 and shows trajectories of the three types of alleles in the first four loci of the averaged genotype. The initial decrease of parameters is soon halted when it hits two alleles per region (*i.e.*, the first four loci). This is a mirror image to the trajectory of [1] as [0] alleles almost do not exist in the region. However, this situation dramatically changes when suddenly [0] alleles emerge after generation 2000. This implies that the parsability bias in communication somehow 'wins' in the locus. When this is visualized, an interesting point becomes clear. The first band in Figure 9.22 (p. 189) shows that the number of parameters in the region is generally larger than that of the original (*cf.* Figure 9.19, p. 187). Also, while the fourth locus is initially occupied by the plastic alleles, it turns out to be occupied by [0] alleles which are disfavored by the parsability bias in the learning process.

This can be thought of a result of niche construction; when a given genic locus in the region is fully occupied by [1], no adaptive difference would be generated

by the bias in communication. No one can learn grammars which include [0] in the first four loci, since no input is available (apart from an occasional mutants). Therefore, everyone is equally (un)fit in the communicative process regarding the locus. However, as long as the plastic alleles occupy the locus, there is a chance that some may learn grammars including [0] in the first four loci, although the chance would be very small. If two agents who are neighbors of each other learn this class of grammars (*i.e.*, grammars which have [0] in the first four loci), then they have a good chance of increasing their fitness out of others (since both learning and communication are spatially bounded). This newly created niche would be further supported by the communication process in later generations, although for such a grammar to be successfully inherited, it has to get through the learning bias in the following generations. Therefore, apart from the early stages where different types of alleles coexist in the same locus in the genepool, if a particular region is canalized to [0], it is most likely from [?] alleles, but not [1] alleles.

Finally, the averaged results of 100 runs are shown in Figure 9.23 (p. 189) and Figure 9.24 (p. 190). All runs are conducted up to 10000 generations. In both figures the biased regions are focused and other regions are omitted. In the result of positive correlation, [0] never appears, while [1] alleles are steadily substituting the plastic alleles. On the other hand, in the decorrelated simulation, the number of [0] alleles slowly but firmly emerges towards the end of the simulations.

From these results, it should be clear that G-P decorrelation by discrepant optimization demands disturbs the Baldwin effect. However, interestingly, as was shown in the results given above, increase of ineluctability itself takes place. Therefore, different from G-P decorrelation by epistatic relationship where plasticity often permanently remains, in this type of decorrelation, reduction of plasticity takes place. However this is not equal to the canalization process as such reductions are not what selection favors. In this regard, this type of G-P decorrelation discloses an interesting aspect of canalization; under a decorrelated circumstance, reduction of plasticity itself is not necessarily equal to canalization, which is favored by natural selection.

9.4 G-P Decorrelation by Complete Separation

In the previous sections, G-P decorrelation is considered within a certain bounded domain. However, as discussed in Chapter 6, such decorrelations are not the only way to implement G-P decorrelation. If the genes rest on sufficiently different loci from that of the fixed genes (but they are 'functionally' related), the canalizing process will be practically blocked. This is because genetic operations on the plastic genes cannot affect the loci of the genes responsible for the innately predisposed

trait. In the conventional Baldwin effect, the canalization process is possible because the plastic genes and the fixed genes are in an *allelic* relationship. If the two genes are separated but occupying a similar region in the same genotype, the G-P relation is, to some extent, decorrelated. However, in this case, it is possible that genetic operations make some correlated behavior in the tier of genotype and that of phenotype, since there are possibilities that hitchhiking effects may take place. Thus, in contrast to high K where the hitchhiking effect reduces evolvability, complete separation of genetic foundations disables evolvability precisely because the hitchhiking effect cannot connect these two.

However, through the niche construction process, BNC can fill the gap. This means that the functionally related, yet genetically separated traits can coevolve through niche construction, and subsequently a canalization process would take place. To test this assumption, further simulations are designed based on Kirby & Hurford.

9.4.1 The Model

Here the important parts of the model are described:

1. Agents
   (a) The Population Size
   The number of individuals in the population at any given time is 200.
   (b) Genotype and Genes
   A genotype consists of 12 genes. A gene can be occupied by one of the two types of allele, namely $\boxed{0}$ and $\boxed{1}$.
   (c) Grammar
   A grammar is represented by 12 alleles –same as the genotype. The number of allelic type is three: $\boxed{\emptyset}$ (*NULL*), $\boxed{0}$, and $\boxed{1}$. $\boxed{\emptyset}$ designates that there is no information on the given part of a grammar. Therefore, this *NULL* allele does not contribute to learning nor communication. All individuals in any generation start with this null allele in all their grammatical strings (irrespective of their genotypes).
   (d) Cognitive Capacity
   An agent has a cognitive capacity which can be used for both *linguistic acquisition* and *linguistic innovation* described below. The size of the cognitive capacity is described as the number of units. Each individual has five units initially. No evolution is involved in this cognitive ability.
2. Spatial Organization
   Spatial organization is the same as Kirby & Hurford. This distribution is used for both learning and communication in the same way as in Kirby &

Hurford.

3. Learning

   The basic algorithm used in this model is the same as Kirby & Hurford (*i.e.*, *m*TLA). Also the spatial organization is used in the same manner. However, there are some significant differences. The detailed procedure is given below:

   (a) If the number of learning trials has not reached the critical period yet, do the following process. Otherwise, finish learning.
   (b) If a given locus of a grammar which is stored in the arena of use is not [∅] (*i.e.*, [0]/[1]), then it is considered to be a trigger (*i.e.*, input), and proceed to (c). If it is [∅], increase the number of learning trial.
   (c) If the number of units of the cognitive capacity is more than one, then compare the trigger with the agent's grammar.
      i. If the corresponding position of the grammar has [∅] (*i.e.*, the locus has not received any input previously), compare the corresponding locus of the agent's genotype (otherwise, proceed to ii). If it accepts the trigger, copy the value to the corresponding position in the grammar. If it does not accept it, then copy the opposite value to the grammar and increase the number of learning trial, and subtract a unit of the cognitive capacity.
      ii. If the agent has grammatical information in the corresponding position in his grammar, compare its value with the incoming trigger. Do the same procedure described above with the grammar (but not the genotype).
   (d) Increase the number of learning trials by one.
   (e) Repeat these procedures until the critical period is reached.

4. The Critical Period

   In the simulation, the critical period (*i.e.*, the number of learning trial) is set to 40.

5. Linguistic Innovation

   After finishing the learning process, check the number of units of the cognitive capacity. If it is more than one, and if some parts of the grammar have [∅], then with 50% chance, pick a point randomly which has no information, and randomly replace the value with either [0] or [1]. Subtract a unit of the cognitive resource. Repeat this procedure until all the units are used up, or all [∅]s are erased from the grammar.

6. Arena of Use

   Copy the agents' entire grammar including [∅] into the arena. This becomes

the next generation's inputs. Note that the initial generation receive no input[5]. This is different from other simulations where random inputs are available for the first generation. So, for the first generation, no learning takes place at all.

7. Communication & Fitness Function
   In the same manner as other simulations of language evolution replicated here, fitness is determined by communicability. The basic procedure is somewhere between Turkel and Kirby & Hurford. The spatial organization is used here too; two adjacent individuals become a speaker and a hearer. Communicative success is calculated based on the Hamming distance of speaker's and hearer's grammars. The fitness function is given as follows:
   $$FITNESS = 13 - N$$
   where N designates the Hamming distance. However, when the two grammars have loci whose both alleles are $\boxed{\emptyset}$, increment the distance. Both the speaker and the hearer are rewarded. Since a speaker also becomes a hearer once, in total, the maximum value of fitness is 26.
8. Reproduction
   Reproduction is exactly the same as Kirby & Hurford; based on roulette-wheel selection, two selected individuals reproduce two children. In the first configuration, genetic operations are introduced: With a probability of 0.001 (*i.e.*, one in 1000 alleles) mutations takes place; one-point crossovers are obligatory in this simulation. In the second configuration, no genetic operations takes place at all. Everything else remains the same.

It should be clear from the above description that the model implements BNC without assuming that the ability of learning and innately predisposed linguistic knowledge are in allelic relationship. In this model, innately predisposed knowledge could be assumed to be either *spandrels* of other abilities which have evolved before, or a consequence of neutral evolution. By any means, such knowledge is neutral in the initial generation. It should be considered that through innovation by the cognitive resource, such hidden traits are 'assimilatorily' exapted. In this regard, the model gives a minimal modification from Kirby & Hurford, while the basic assumptions behind the model are quite different. Also, learning in this model should not be captured as the learning ability for language acquisition *par excellence*. Rather it should be thought of as a special application of a domain-general learning ability.

9.4.2 Results

The results are shown in the figures (from Figure 9.25 to Figure 9.30, pp. 195-198). First, Figure 9.25 shows the overall result of the evolution. The number of [0] and [1] in this averaged grammar rapidly increases in the first 100 generations. Since for the very first generation, no linguistic input is available, no learning takes place. Subsequently, the whole residual cognitive resource can be used for linguistic innovation. Based on such 'innovated' languages, the first generation attempts to establish communications. Thus, in the next generation, some linguistic inputs are available for learning. However, as the arena of use is not yet sufficient for covering a whole grammar (with the five full units of the residual cognitive resource, on average, only 2.5 bits of grammar can be innovated), another innovation process takes place in this generation. As this cycle is iterated, eventually the arena of use is saturated.

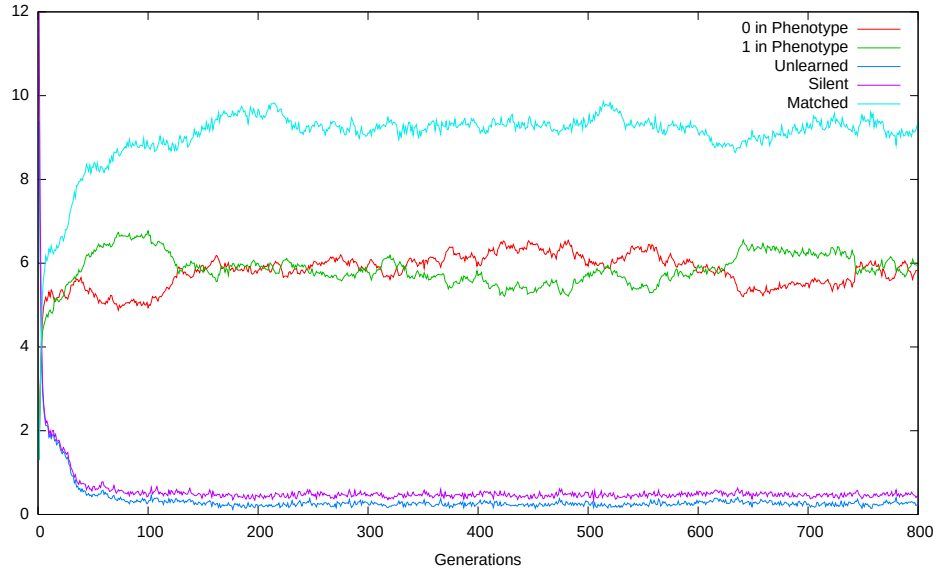

Figure 9.25: G-P Complete Separation 1

This iterated process goes hand in hand with a canalization process of such learning. As the amount of the cognitive resource is capped, it barely covers less than a half of the whole grammar (five full units of the cognitive capacity compared to 12 bits of grammatical information) is covered. Thus to increase the size of the grammar, some parts of the linguistic knowledge have to be canalized so that such parts do not have to use up the limited ability.

In Figure 9.25, the line keyed as "Unlearned" shows the number of bits in the averaged grammar which do not receive any input. There are two reasons that can be thought of for this unlearned situation. First, no input comes from the arena of

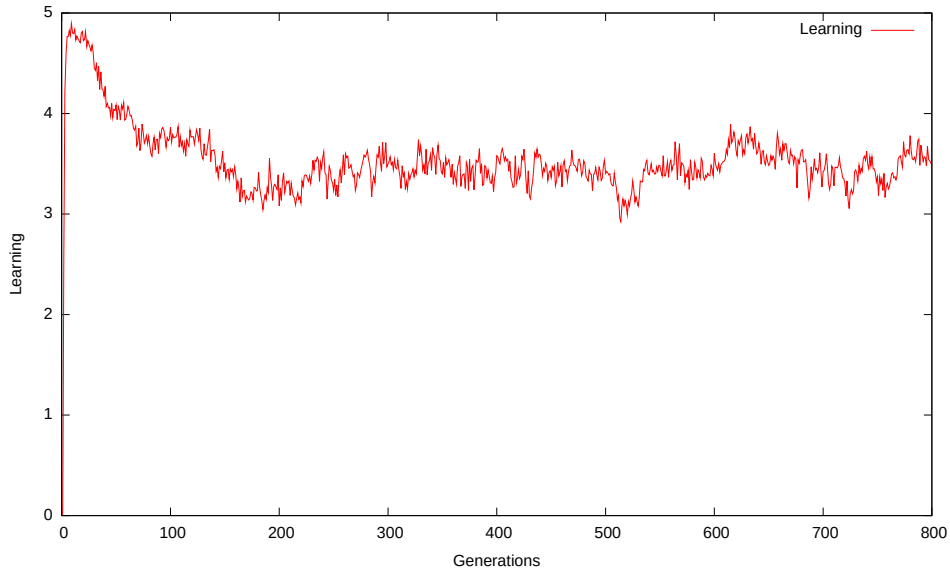

Figure 9.26: G-P Complete Separation 1 –Learning

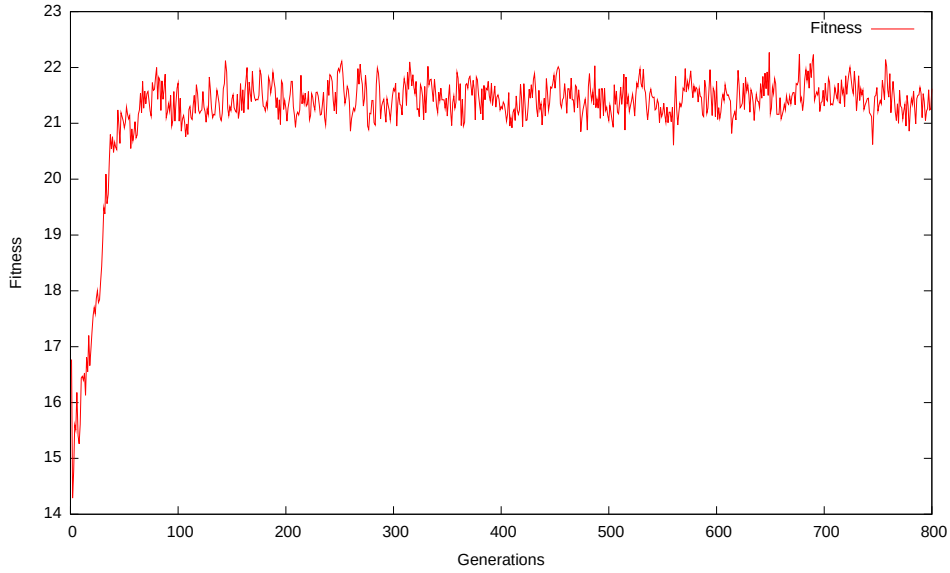

Figure 9.27: G-P Complete Separation 1 –Fitness

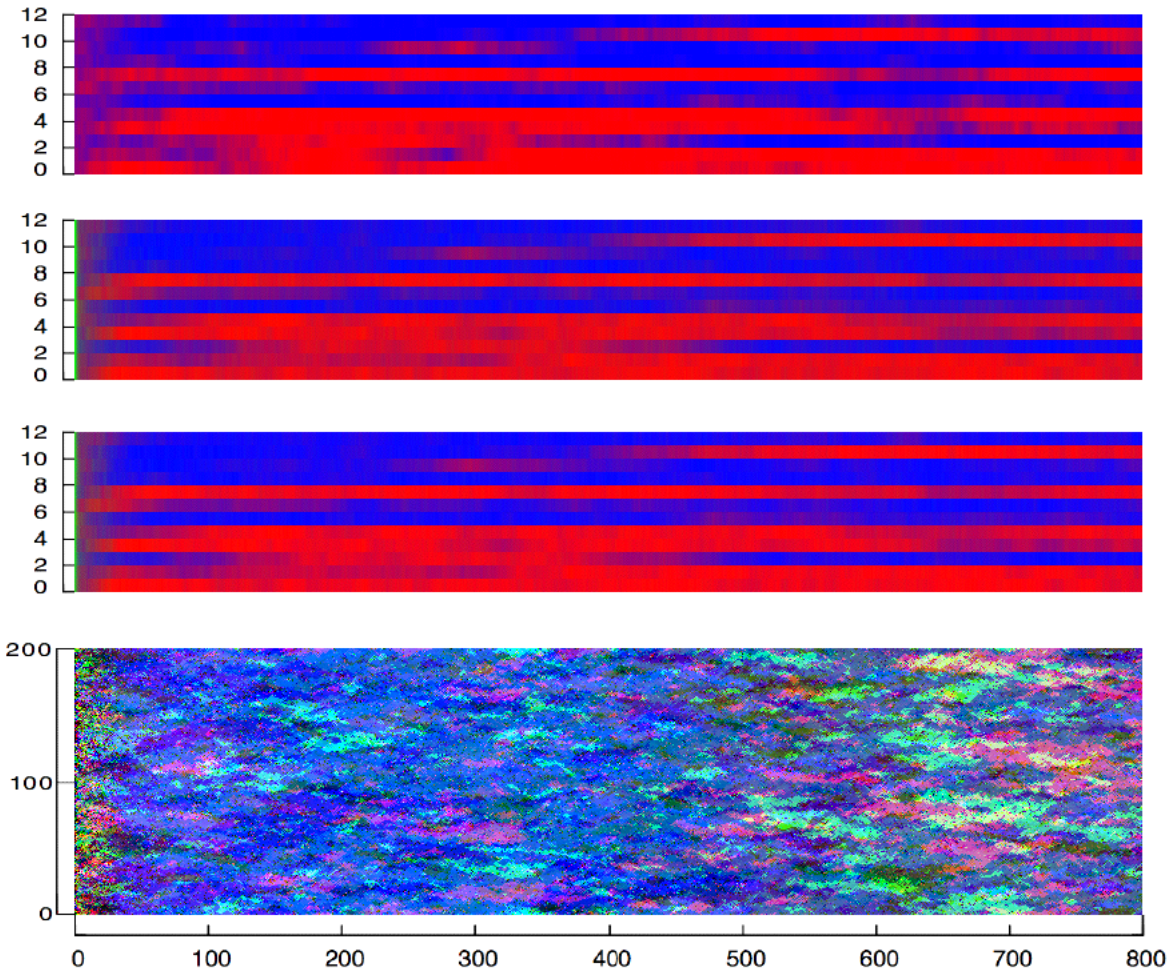

Figure 9.28: Spatiotemporal Graph of G-P Complete Separation 1

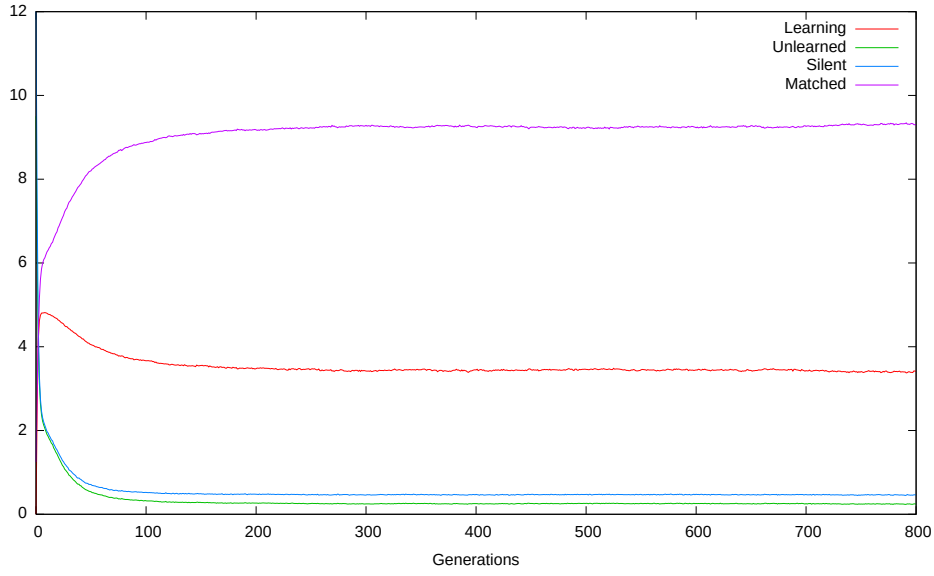

Figure 9.29: The Averaged Result of 100 Runs

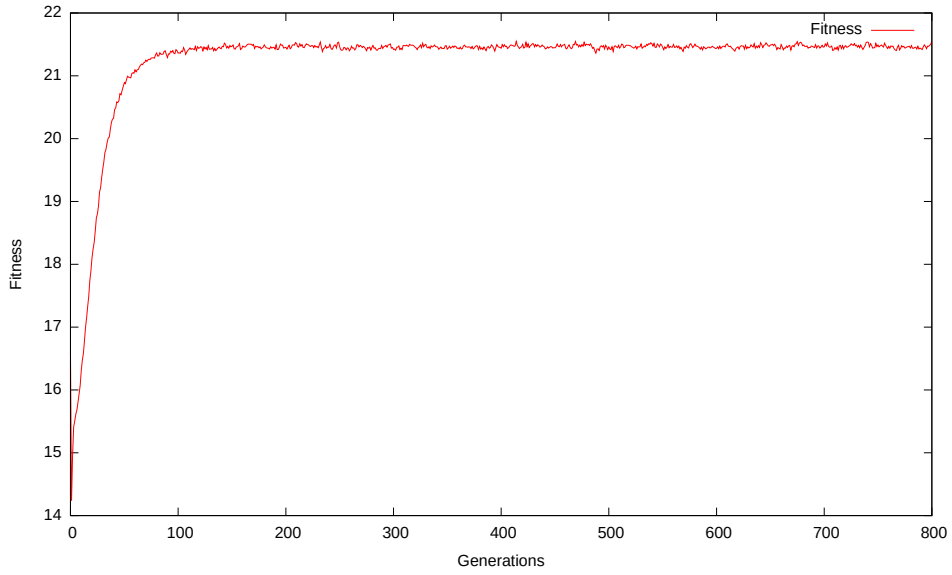

Figure 9.30: The Averaged Result of 100 Runs –Fitness

use for the position. Alternatively, if learning is used up before any input comes in, and the input and the genetic information on the corresponding locus are different, the position on the grammar remains unlearned. Such unlearned positions in the grammar fade rapidly as innovations produce somewhat *ad hoc* inputs. "Silent" designates positions where $\boxed{\emptyset}$ alleles occupy. Thus such parts are totally *silent* for the grammar. Although such silent parts would never be zero, the number of such parts is well below 0.5 per grammar.

Figure 9.26 (p. 196) nicely shows how learning behaves during the early stage of the simulation. The figure shows the number of learning trials which requires the cognitive capacity at the end of the learning process (but before the innovation process). It begins from zero as the initial generation cannot get any input from the 'previous' generation. However, since the generation can use the resource fully for innovation, from the next generation, the number of such trials surges up to the near-maximum. However, within less than 50 generations, it starts decreasing; through the canalization process, the cognitive capacity is 'freed' from language acquisition. Roughly around 200 generations, this canalization process is finished and the curve reaches a plateau.

In Figure 9.27 (p. 196), the averaged fitness is shown. Fitness quickly increases to the sub-optimal level. The pace of evolution is quicker than other aspects of the simulation. Most notably, compared to the evolution of consumption of the cognitive capacity, which shows a comparatively slow evolutionary process, fitness does reach the stable condition a lot quicker. Since the canalization process is

primarily triggered by costs in fitness, this discrepancy poses an intriguing question. However, in this thesis, we will not address this point.

Figure 9.28 (p. 197) gives graphical representations of evolution. The first stripe is the representation of the averaged genotype. The second is the averaged grammar after the learning process. The third is the grammar after the innovation process takes place. As is apparent in the stripes, all the three stripes quickly become almost identical. This supports the above analysis that the canalization effectively takes place. Also, the very high similarity of the second and the third indicates that innovations do not add any significant impact on the final state of the grammar (the thin green band in the second and the third indicates regions where some of the grammar is unlearned and/or silent).

As the spatiotemporal figure shows (in Figure 9.28), languages are spatially organized, although comparatively high noises are observable. Finally, the averaged result of 100 runs are shown in Figure 9.29 (p. 197) and Figure 9.30 (p. 198). These assure that the above result is a typical case of this simulation.

The result of the second configuration (*i.e.*, the 'no-mutation-no-recombination' configuration) appears from Figure 9.31 to Figure 9.34 (pp. 200-201). Somewhat surprisingly, the result is almost indistinguishable from the first configuration. Through the exaptation process, previously hidden genetic variance is put through the canalization process: This nicely appears in both Figure 9.33 and Figure 9.34; as the average number of learning process shows, through the canalization process, the cognitive capacity is unloaded from the learning process. The top three bands in Figure 9.34 confirm this analysis; compare the first and the second & the third. The genotypic information is properly reflected on their grammars. The averaged results (Figure 9.35 & Figure 9.36, p. 202) also confirm that in this mode of evolution, genetic operations have a somewhat ancillary role.

The above simulations successfully demonstrate that even under a complete G-P separation, BNC can make the population evolve. Especially, as the high matching rate between grammars and their corresponding genotypes shows, the canalization effect seems to takes place properly. Moreover, as is apparent in the results of the second configuration, in this mode of evolution, genetic operations are essentially unnecessarily. Together with the results of the simulations in Section 9.2.4, this indicates that this mode of evolution is fundamentally resistant to epistatic configurations.

In the following simulation, to check whether learning really induces this exaptation process, the concept of linguistic biases given by Kirby & Hurford is introduced. Based on the simulation in Section 9.1.2, the parsability bias is introduced in the learning process (but not in the communication process, as it is irrelevant for the

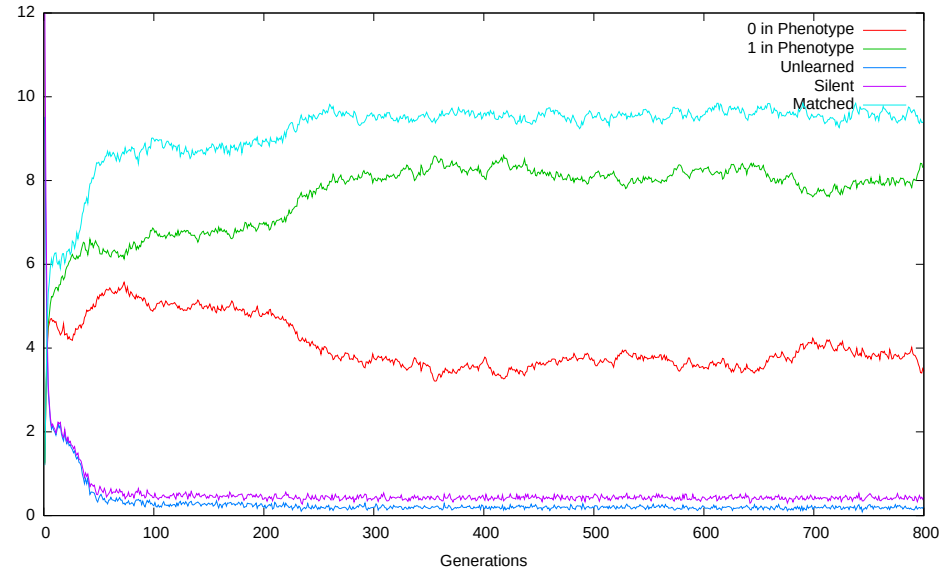

Figure 9.31: G-P Complete Separation 2

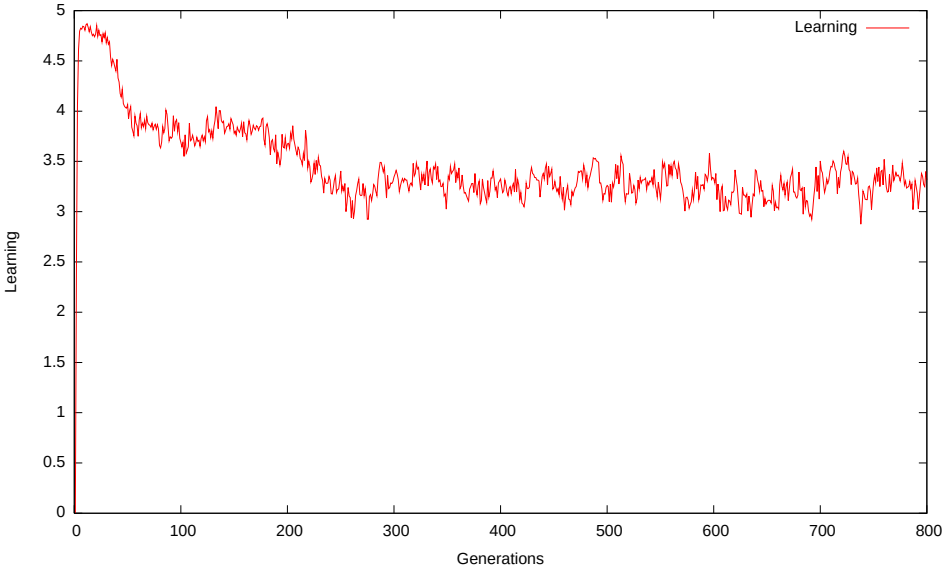

Figure 9.32: G-P Complete Separation 2 –Learning

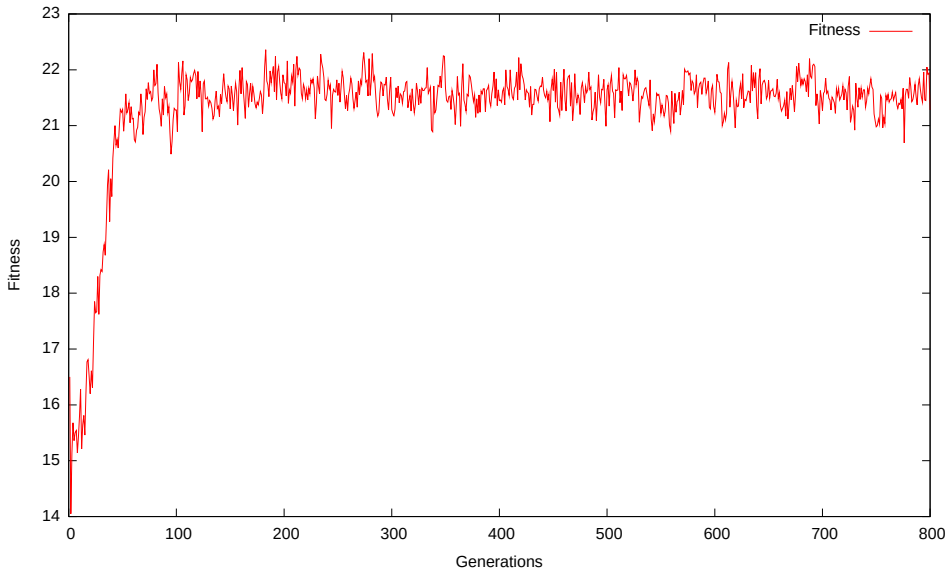

Figure 9.33: G-P Complete Separation 2 –Fitness

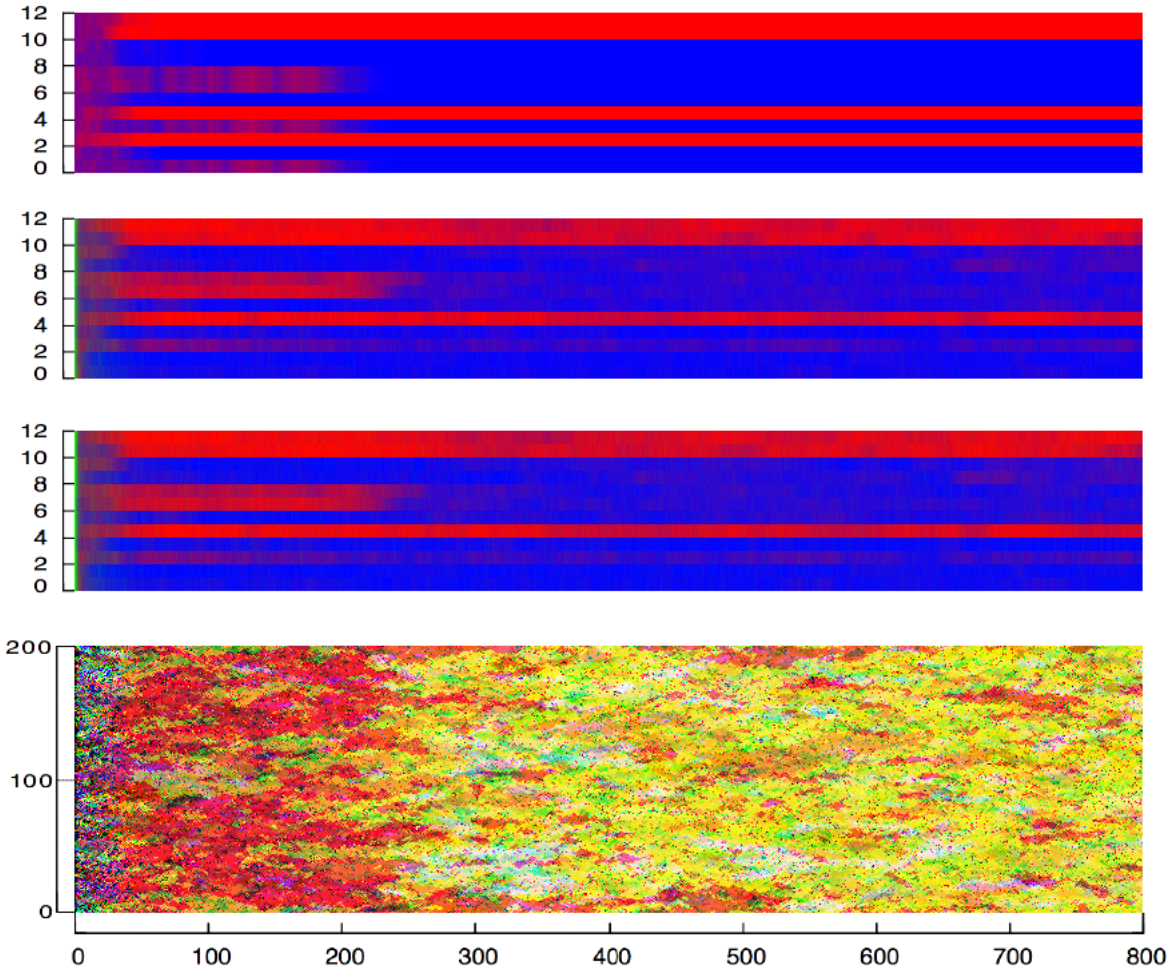

Figure 9.34: Spatiotemporal Graph of G-P Complete Separation 2

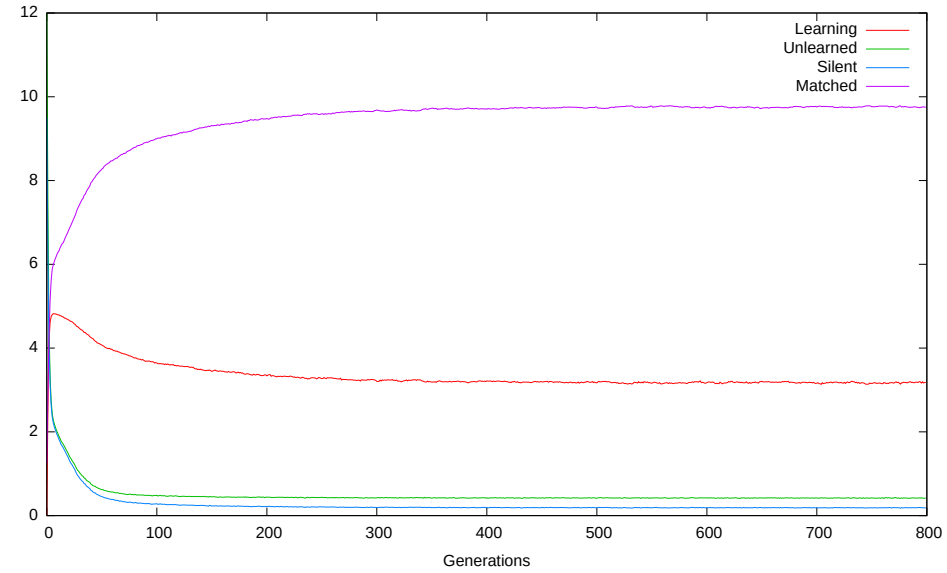

Figure 9.35: The Average Result of 100 Runs

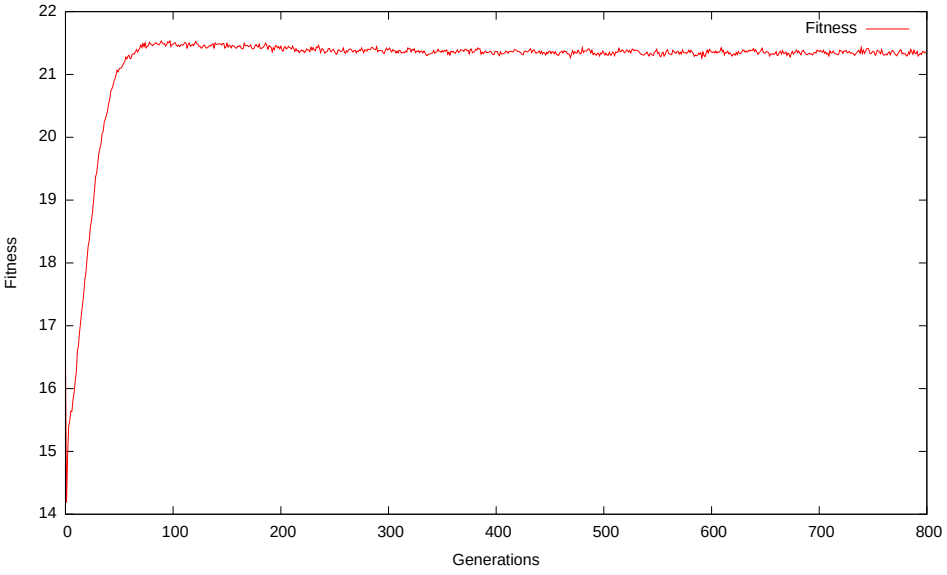

Figure 9.36: The Averaged Result of 100 Runs –Fitness

focal topic). In the same way as Kirby & Hurford, with 10% chance, the learning process prefers increasing the number of [1] in the first four loci of a grammar. The result is shown in Figure 9.37 (p. 203). The figure shows the number of the learning trials which requires the cognitive capacity, the number of [1] alleles in the first four loci in the averaged genotype, that of the averaged grammar, and the amount of discrepancy between the regarding loci[6].

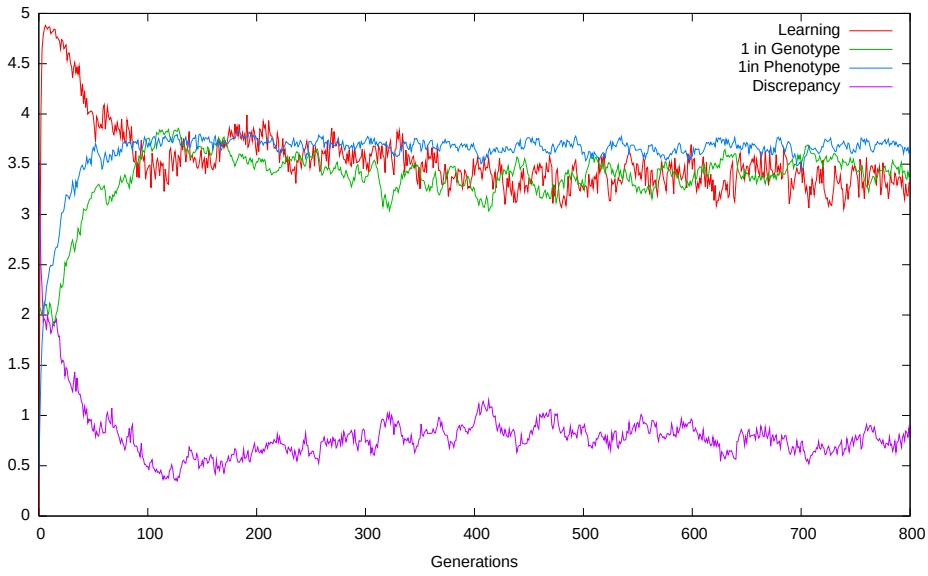

Figure 9.37: Biased G-P Complete Separation

The number of learning trials evolves almost the same as that of the original configuration. However, the number of [1] in the first four bits of both genotype and grammar is higher than the expected value of the non-biased configuration (*i.e.*, 2). This means that learning successfully biases grammars to be equipped with [1] in the first four bits. Also, this is successfully transmitted to the genotype through the exaptation process. The spatiotemporal figure (Figure 9.38, p. 204) assures this; the first four bits of the averaged grammar (both after the learning process and after the innovation process) are almost fully occupied by blue (*i.e.*, [1]). The corresponding regions in the genotype are also mostly blue. However in the genotype, some red regions are found in the first four bits. However learning properly overturns the color (compare the red region with the following two figures).

The averaged result of 100 runs also confirms the above result (Figure 9.39).

[6]Genotypes and their corresponding grammars are compared on the focal region. If values are different, it is counted as discrepancy.

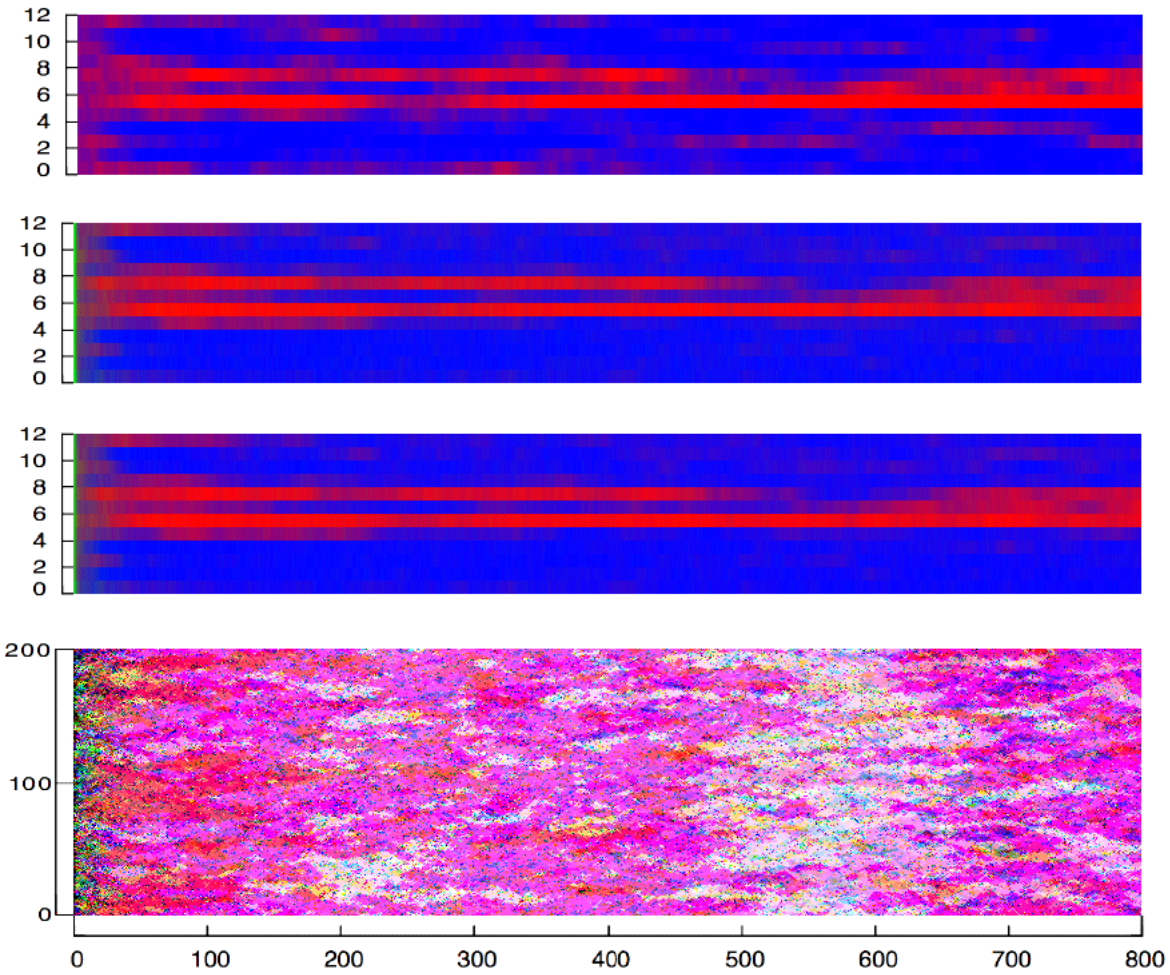

Figure 9.38: Spatiotemporal Graph of Biased G-P Complete Separation

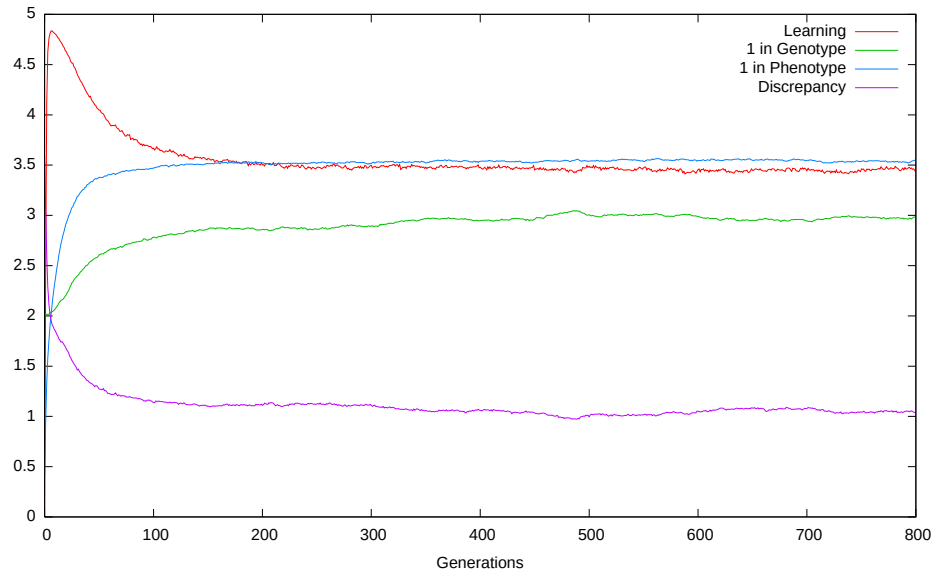

Figure 9.39: The Averaged Result of 100 Runs

9.4.3 Discussion

From these results, it should clear that BNC is highly effective even if the genetic foundation of learning and grammatical knowledge are remotely distant; the cognitive capacity is initially used for a linguistic innovation so that it produces somewhat *ad hoc* grammars. Based on such grammars, later generations' linguistic learning takes place. Once the amount of linguistic input reaches some degree, linguistic learning becomes a burden. This pressure makes the canalizing effect take place. Although the simulation is simple, it sufficiently demonstrates a cyclic application of linguistic innovation and canalization.

Note that, in this model, the cognitive capacity plays three important roles. First, it allows agents to express information in their grammars. Secondly, it allows modification of the information based on linguistic inputs. And finally it allows the agents to innovate new linguistic expressions. This innovation process is only allowed when the capacity is unloaded from language learning. If no sufficient triggers are available from the input data, the cognitive capacity can be preserved for the innovation. Also, if the genetic information of a grammar matches to the given linguistic environment, the burden of learning is also eased.

This multi-faceted aspect of the cognitive capacity makes it as a more domain-general like ability. While, as this simulation only concerns language evolution, this aspect of the cognitive capacity is not focused in this thesis, the multiple roles of the capacity are conceivable as a reflection of the aspect. The complete genetic separation of learning and linguistic knowledge supports this view. As they have completely different genetic bases, it could be thought of as no domain-bounded relationship between these two functionally similar properties.

Although in the simulation the cognitive capacity is not genetically represented, this is merely to implement the separation. It is perfectly possible to design the simulation to have a genetic basis of the capacity whose evolution is not correlated with that of linguistic knowledge, however, as this would not affect the result of the simulation, we omit it here.

9.5 Summary

In this chapter, various simulations have shown that the conventional Baldwin effect which stipulates G-P correlation is effectively blocked by different types of epistasis. This brings a serious consideration of feasibility regarding the G-P correlation model in the Baldwin effect.

However, communications inherently involve a niche construction process, the simulations of language evolution based on Kirby & Hurford still exhibit some degree of the canalizing effect through BNC. Nevertheless, as the degree of epistasis

increases, the dependency among different genes in a genotype also increases. Consequently. genetic disturbances become overwhelmingly strong. This is the reason that with high K values, the canalizing effect disappears.

This result is overturned when such genetic reshuffling is excluded. Through the pure exaptation process triggered by niche construction, some of the previously non-adaptive genotypes become adaptive. This process is highly expedited compared to 'with reshuffling'. Presumably, this is because a small number of genotypes becomes strongly adaptive within the population, and they quickly prevail.

In the last section, in lieu of epistatic G-P decorrelation, complete separation of learning and canalized linguistic knowledge is introduced. With this separation, it is more easy to conceive that learning is a domain-general ability. The learning allows for both acquiring linguistic information and innovating new expressions. Increasing the amount of information through innovation is potentially adaptive, but to do so, an individual has to increase properly canalized genes in her genotype so that the cognitive capacity can be spared for the innovation. Through these complex interactions of learning, innovation, exaptation, and canalization, genotypes which well match to the extant languages are rapidly selected. As a result, a large part of a given grammar is canalized. This also gives an extra space for the cognitive capacity.

Finally, although the simulation is bounded to a certain degree; in a real world, such a freed cognitive capacity could be used for yet another evolutionary process of language.

Chapter 10

Discussion

In the last chapter, the computer simulations show that the conventional mechanism is indeed susceptible to G-P decorrelation; if the value of K increases (*i.e.*, the level of decorrelation increases), the degree of the Baldwin canalizing effect weakens. However, under the BNC mechanism, this decorrelation is logically irrelevant, especially if a given genepool contains a sufficient genetic diversity. This is proven by the later simulations.

These simulations are simple and some assumptions are admittedly crude, yet the data available from them sufficiently support the robustness of BNC. As discussed in Chapter 6, BNC does not essentially rely on genetic operations (*i.e.*, mutations and recombinations); through the process of niche construction, some exaptation of previously neutral genes takes place. Some of the simulations show this indeed takes place experimentally.

Given these results, in this chapter, some new perspectives which may influence our future studies of BNC are discussed.

10.1 Assimilative and Dissipative Exaptation

As noted in elsewhere, in his recent literatures (*e.g.*, Deacon 2003), Deacon has developed his account of coevolutionary theory of language and brain described in *The Symbolic Species* (Deacon 1997). In particular, based on the concept of niche construction, he deployed a similar theory to BNC. However, there is at least one significant difference between BNC and his new theory.

Basically by reviewing the ambiguous usage of the Baldwin effect in the theory, Deacon provides a logically-more-sophisticated evolutionary account of dissipative allocations of linguistic abilities. Recall that in *The Symbolic Species*, he discusses that linguistic abilities are implemented in a constellation of cognitive (and physiological) abilities. Crucially, Deacon considers that evolutionary processes involved in language evolution do not work for increasing innately predisposed linguistic abilities, but for 'decreasing' contributions of such predisposed abilities. Therefore,

his evolutionary account of language flows in a reverse direction of the standard Baldwinian accounts.

In the recent literatures, to properly capture the logical flow of the dissipative evolutionary process, Deacon terms the process "the *reverse* Baldwin effect"; because the process decreases innate attribution rather than increases it. New factors to create this effect are called the "*masking*" and "*unmasking*" effects. Deacon argues that first innate predispositions were '*de*-differentiated' each other because learning begins to play a role; a type of phenotypic neutrality emerges. Avital & Jablonka (2000) also consider a very similar concept:

> Since plasticity of higher animals can mask both environmental and genetic variations, many genetic variations are protected from selective elimination and can accumulate. The net effect is a large reservoir of genetic variation underlying the organization of the nervous system. This variation is exposed and recruited when the environment changes.
> (Avital & Jablonka 2000, p. 323)

Recent studies in biology have revealed that organisms are often equipped with self-regulatory and self-organizing capacities which play a role of compensating for absences of specific genes. This can be considered as a result of the masking effect. Niche construction, he considers, would be attributed to this process.

Once the masking effect takes place, Deacon assumes that degradation of masked abilities is inevitable. This is a natural assumption as genetic drift often takes place under a phenotypically neutral condition. He reckons that this degradation induces the second process, namely the unmasking effect. It is unclear that how such an unmasking effect –a type of exaptation process is brought in. However, importantly this is an exaptation process but it is not assimilative. Rather, the process is dissipative; since the once necessary factor to acquire (or invent) a given ability has been degraded through the masking effect, a constellation of different abilities now have to play the role cooperatively. Thus the unmasking effect is thought to induce "*highly distributed parallel synergistic consequences –with the potential to significantly amplify adaptations*" (Deacon 2003, pp. 95-96). Deacon reinterprets Waddington' works of genetic assimilation with this concept:

> Waddington implicitly attributed genetic assimilation to the unmasking of variants, otherwise unexpressed, by the introduction of new selection pressures in the form of environmental stresses.
> (Deacon 2003, p. 96)

He considers that Waddington's experiments on the fruit flies are fundamentally equivalent to an unmasking process; what are revealed by the changed environmental factor(s) (*e.g.*, *Bithorax* phenocopies) are not attributed to a single cause (*e.g.*, a gene responsible for the phenotypic reaction), but a variety of different causal factors which are scattered differently in different individuals. By inbreeding individuals who express such a phenotypic trait, Deacon argues, Waddington successfully enhanced the synergistic effect of such factors, so that the phenotypic trait 'ineluctably' emerges even without the given environmental factor(s). The source of cost would be, in this case, attributed to differences regarding ineluctability of expressing such a trait; if multiple factors collaboratively express the trait more ineluctably than a single cause does, the difference of such stability will be the cost.

As Deacon has his own specific linguistic theory in his mind, the process of the masking effect regarding language evolution is discussed within this framework; the reverse Baldwin effect can have enhancement of a linguistic ability through a niche construction process, while nothing becomes more innate.

As an example the masking & unmasking evolutionary process, Deacon provides a case of the evolutionary relationship of 'frugivory' and the endogenous synthesizing ability of vitamin C in fruit-eating primates. Fruit-eating primates have known for their lacking ability of synthesizing vitamin C. This is because the gene for the final enzyme for synthesizing vitamin C has been degraded for those animals. Deacon suspects that the ubiquity of vitamin C available in fruit is attributed to this evolutionary degradation process; high availability of fruits masked the importance of the synthesizing ability, and as a result a degradation process took place. Instead, the ability to find edible (*i.e.*, sufficiently ripe) fruits became important. To attain this, fruit-eating primates have developed not innately predisposed ability to find suitable fruits, but a set of extant abilities which collaboratively work for finding such resources. A simulation conducted by Wiles *et al.* (2002) shows this would be the case.

We consider that this provides a new perspective on BNC. Recall that in Chapter 6, we succinctly discuss that niche construction is a type of process which creates STEs. This is especially true for internal cooperative niche constructions, as such a mode of niche construction often induces environmental equilibria. Then Deacon's idea of dissipation is a highly informative take-home message.

The key point is that once STEs are created, individuals could rely on learning, as the inputs necessary for the particular learning are stable under such an STE. While the canalization process is a narrowing process of a reaction norm, creating an STE is a fixation process of a particular environmental condition. Usually, such

a fixation is given by nature, but not by organisms. However, because niche construction is an organism-referent process, the fixation process by niche construction is ultimately created by the population. Once the environmental condition is fixated, a development of a certain trait becomes highly ineluctable. Having said that, it is still possible that even within the stable environment, some improvement of ineluctability would take place. After all, nothing can be perfectly deterministic; even within a highly environment, some uncertainty may exist on a given development. To increase ineluctability, increasing the contribution of learning is logically plausible. This 'within-STE' ineluctability improvement by learning may make the canalization process redundant. However, as the experiments demonstrate, if the learning capacity is somehow capped, to proceed the assimilate-stretch process, the canalization effect may take place to replace learning; this would be a cost of learning in this type of process. These problems will be addressed in future studies.

10.2 Language as a Complex Dynamic Adaptive System

For a decade after Pinker & Bloom published their seminal paper (Pinker & Bloom 1990), the majority of studies in language evolution have been devoted to the biological aspect. However, in recent years, studies in the cultural aspect of language evolution have begun to provide intriguing results. For example, Morten Christiansen and his colleagues have shown that some of the allegedly 'non-functional' aspects of linguistic knowledge can be both cognitively and evolutionarily accountable. It is known, for instance, that a linguistic constraint called "*subjacency*" Exhibits a strongly dysfunctional aspect. Thus, it has been reckoned that functional explanations cannot be available for at least a core part of linguistic knowledge. However, Ellefson & Christiansen (2000*a*, 2000*b*) show that such a constraint may evolutionarily emerge due to limitations on sequential learning during language acquisition.

Kirby and his colleagues have also been working on a possible aspect of language evolution from the cultural evolution perspective (Kirby 2000, Kirby 2001, Kirby 2002, Kirby & Hurford 2002). They invent a very minimal model of cultural evolution called the "*Iterated Learning Model*" (**ILM**); essentially it does not contain population nor genetic representation; simply a sequence of learning and teaching exists. More precisely, in a given world, two agents always exist. One is a learner and the other is the teacher. When a learner becomes an adult, she becomes a teacher of the next 'generation'. What they learn is a system of meaning-symbol mappings. As they do not contain genes, what they have learnt cannot be inherited genetically. Instead, languages (*i.e.*, the meaning-symbol mapping systems) are passed through learning. Therefore, no explicit distinction between underlying structures and manifested entities exists. Possible styles of meaning-symbol

mapping systems are various; some may be completely random and others may be systematic. With some learning algorithms, when the total amount of data a learner can get is large, random mappings often emerge; with a sufficient amount of input data, learners can reliably learn random mappings. However, when the size of the window is narrowed, interestingly, more systematic mapping systems emerge. Such mappings are compositional. The logic behind this is because the amount of inputs a learner can gain is small, random mappings are no more sufficiently learnable. On the other hand, compositional mappings are systematic, so that with a small amount of data, a comparatively large part of the mapping system can be covered.

In both studies, a language is implemented as a complex dynamic adaptive system. It changes its style over time so that it fits to learners acquisition capacity. Therefore, from this view, one may perceive that it is language that adapts to human cognitive capacities. This is an attractive view at least for the following three reasons:

First, it may provide rather direct evolutionary explanations for long standing riddles of modern linguistics. Learning plays a crucial role in this type of study –it makes language as both the underlying structure and the manifested entity. As Elman *et al.* (1996) consider, learning may well have a key role in language universality. According to them, learning is one ability which is consistent across different environments. Therefore if something constraints the language acquisition process, it would be reflected in the end-product, namely individual languages. Also, in a similar vein, as such learning ability is ubiquitous, constraints on learning provide causal explanations in language universality.

Secondly, by considering cultural evolution, one can avoid (or at least ease) the 'adaptiveness' concern. In genetic evolutionary theories, one has to consider the effect on reproductive success regarding a concerning trait (in our case, language). As in this thesis, we have considered communicative success through linguistic communications. Although it is almost a banal truism that success in linguistic communication contributes reproductive success somehow, it is yet highly controversial how exactly such a thing contributes success. However, in cultural evolution, adaptiveness does not often have to be explicitly measured; it is simply a matter of whether or not it is acquired by a learner. If not, it just disappears.

The third reason is the pace of the evolutionary process. This has already been discussed elsewhere, but the process of language evolution has been considered exceedingly rapid. Remember that, one of the reasons the Baldwin effect has gathered attention in this field is its expediting effect. Since in pure cultural evolution, no genetic process is required, its pace is thought be very rapid. Although BNC could

accelerate the evolutionary process to a large extent, that of cultural evolution is thought of as even faster.

In this thesis, we have not captured language as a complex dynamic adaptive system. Fundamentally, if language is dynamically adaptive so that it conforms to the human learning capacity, it would be possible to consider that the canalizing effect on BNC would be weakened. Together with the discussion in the above section, this property may provide a new avenue in the study of BNC. To investigate this avenue, we have to elaborate our simulation models; our models are based on simple implementation of the P&P theory. However, as in generative linguistics, actual representations of such principles and parameters are largely undetermined yet. That is, although the theory provides a framework of language acquisition, it does not specify the nature of each principle/parameter. This blocks us from considering more specific representations of input data in the models. As the model of BNC is not necessarily bounded by this representation of language acquisition, our implementation of language and the language acquisition device should be elaborated.

Appendix A

The intriguing point BNC is that since niche construction is a dynamic process provided by learning, it provides a mode of evolution which is different from what genetic operations provide; niche construction makes organisms evolve not by providing genetic diversity (and consequently phenotypic diversity) to the population, but by exposing hidden genetic diversity by modifying environmental factors; this triggers natural selection. This opens the way to consider language evolution as a case of exaptation. However, for BNC to work under this type of circumstance, an important condition has to be met. That is, a given population has to have a decent degree of genetic diversity. Since if the relationship between genotype and phenotype is decorrelated, genetic operations are mostly useless, the initial diversity plays a key role. For example, if such a diversity is low, or biased, necessary genotypes (ones which would become adaptive on a particular niche) may well not be available. In this case, sufficient canalization would not follow.

However, it is somewhat obvious that niche construction itself cannot create genetic diversity; it is an exclusive feature of genetic operators. Then a pertinent question is that how such genetic diversity is created *before* a particular niche construction takes place. By and large, two different types of causes are conceivable; the first is neutral evolution and the other is evolutionary spandrels. In this chapter, some backgrounds to these phenomena are considered. Although the discussion in the following passages will not be experimentally supported and hence is purely on an argument-basis, it is expected that this will shed some basic light on further study of this field.

A.1 Neutrality in Evolution

A.1.1 Genetic Variation under Genotypic and Phenotypic Redundancy

The two primary factors in biological evolution are heritable variation and selection. A trait can evolve if the following criterion is met: A population embraces some degree of heritable variation of the trait among its members, and an environment surrounding the population can 'distinguish' the variations. The definition of the term "distinguish" is that the environment affects the reproductive success which

consequently modifies the frequency of the variations in the next population. By this mechanism, the population changes its shape in response to the environment. Although the heritable variations and selection may not be a sufficient condition, they are *sine qua non* of evolution. These two factors are the necessary conditions not only for biological evolution, but also for other types of evolution. For example, GA utilizes this very feature; the minimum requirements of an algorithm which can be called a GA are these two factors. GAs have proved that the combination of heritable variations and selection is indeed the primary engine of developing some traits or aspects of an abstract phenomenon.

The source of heritable variations in nature is primarily attributed to genetic mutations. Genes are responsible for inheritance of traits in individuals. Reproduction of offspring can be ultimately grasped as a special case of gene replication. With a certain probability, during the replication process a new variant of a gene is copied from a gene. This 'miscopying' of a gene during the reproduction process is generally thought of as the main source of heritable variations in the population. Also, mutation is crucial in GAs. Evolution both in nature and *in silico* is enabled by heritable variations which are triggered by genetic mutation. Variations triggered by genetic mutations are, however, not always guiding the population in a good direction; most mutations are strongly deleterious. Together with the very minor probability of mutation, positive variations emerge even less frequently. Thus the pace of evolution is generally very slow.

However, as any fine sieve allows passing minutely different sizes of grains, it is impossible that the selection mechanism picks up (or weeds out) only one from all other variations. In other words, some variations are neutral to a particular environment. If the environment is harsh for the population, that is equal to saying that selection is strong. On the other hand, if the environment is friendly, that means that selection treats a number of variations as the same. As genes are the source of the variations, the selection mechanism cannot differentiate some genes. In other words, some variations are redundant in the face of natural selection. This lack of differentiation (or existence of redundancy) of variations from the selection mechanism is the core notion of 'neutrality' in evolution.

As canalization can be thought of in different levels (Section 2.4.1), there are different levels of neutrality in evolution. That is, organisms are The following descriptions are a rough summary of these redundancies:

Genotypic Redundancy There are a couple of causes that produce genotypic redundancy. First, at the level of DNA, different sequences of nucleotides (called codons) code the same type of amino

acid because the genetic code is redundant. This is likely since three nucleotides from four possible nucleotides form a codon. As the number of types of amino acids is just twenty compared to 4^3 (=64) possible codons, plenty of redundancy exists. From this, it is obvious that if some nucleotide changes occur among such redundant codons, such substitutions are not detectable from the level of amino acid (with some exceptions). This is called synonymous substitution. One of the best examples of this is the codons to the amino acid leucine. The amino acid is coded in six different codons *AAT*, *AAG*, *GAA*, *GAT*, *GAC*, and, *GAG*.

Mutations in introns are another type of neutral mutation at the genic level. All sequences within a gene are not necessarily 'meaningful'. Some sequences are removed after transcription by a process called gene splicing. This type of sequence is termed an "*intron*" while a meaningful sequence is named an "*exon*". Thus these sequences do not code for any amino acid; consequently mutations within introns are neutral.

There is a further case that genetic mutations do not have effects on phenotypes. Pseudogenes are non-functional copies of functional genes. As such genes are not expressed in the phenotype, mutations which occur on the pseudogenes are neutral.

Phenotypic Redundancy There are a number of cases we may consider regarding neutrality in phenotypes. Some cases of phenotypic neutralities are self-evident. As argued above, for instance, minute differences among phenotypes might not be distinguished. Others are, however, not necessarily obvious as above. Environmentally canalized development is one of such examples. Some characters are robust regarding their final phenotypic outcome if a considerable amount of genetic and/or environmental variation exists. Two different genotypes may express different phenotypes in an early stage of the development process. In the course of development, however, such differences in the phenotypes disappear as final phenotypes expressed from the genotypes converge onto one same final phenotype. The same thing can be said for two different environments; phenotypes expressed from one (or more) genotype under different environments still develop into the same final phenotype. Plasticity also contributes to phenotypic redundancy. If

a phenotype is plastic, it is possible that under different environments a given organism still attains the same phenotypic character from different experiences. The same can be true if two different genotypes express plastic phenotypes. Plasticity, however, has the opposite effect too; one initial phenotype might reach two different final phenotypes.

A.1.2 Mutation, Genetic Drift and Random Walk

Genetic Drift and Random Walk

Genetic mutations which do not have an effect on phenotypic values or fitness (*i.e.*, neutral mutations) result in randomly changing allele frequencies where the given mutations occur. This process can be concisely described in "*the genepool model*". This model concerns a profile of the whole population's genetic movement but not individual organisms' evolutionary trajectory. Thus, in the model, the concept of individual is merely a container of specific alleles and the population is described as a mass (*i.e.*, genepool) of such alleles. Breeding is a process where the alleles (from the previous generation) in the mass are drawn and put into new containers (*i.e.*, gametes). Suppose we have a group of individuals whose genotypes are diploid, and have two types of alleles in one locus, a^1 and a^2. Suppose also that all individuals in the population have an equal gene length. Importantly, neither selection nor mutation takes place; mating and breeding are driven by a completely random factor and the original two alleles are never substituted for other alleles (*i.e.*, no mutation). The genotypes of all parents form a genepool from which offspring is produced by breeding.

If the breeding process continues a certain number of generations, then the frequency of a specific allele, say allele a^1, in the particular generation will follow a binomial distribution even if it starts from a different proportion. If all members in the initial population have the same genotype, over a number of generations increasing dispersion of the gene frequency will be observed. This unguided dispersion of gene-frequency is called "*genetic drift*". Genetic drift is often considered a type of random walk in the possible evolutionary search space. Adding mutation but no selection to the model may bring a complication, yet the basic idea will be intact. As genes that do not have an effect on phenotype or fitness can be randomly inherited to next generations, neutral genotypes often become the subject of genetic drift.

Genetic Drift and the Role of Population Size

However, the model described above is effective only when the size of the population is quite small (*e.g.*, 100 breeding pairs or fewer). When the size is large, it starts

to follow *Hardy-Weinberg* Law. Strictly speaking, the law as well as the case of genetic drift is applicable only when the following seven conditions are met:

1. Infinite population size
2. No selection
3. No mutation
4. All members breed
5. Totally random breeding
6. Every member produces the same number of offspring
7. No migration in or out of the population

Effectively, the law is applicable to a population whose size is more than 100 or so where the effect of genetic drift is minimal. Consider a simple case that N (a finite number) alleles are selected from a genepool; the types of the alleles are, again, either a^1 or a^2, each with a frequency of 0.5. If N=10, the frequency of a^1 allele and a^2 allele is unlikely to be 0.5 due to sampling error. This is another way to describe the mechanism of genetic drift. When we increase the number, say to 30, we still observe sampling error, but the proportional deviation (from 0.5) is smaller than the case of 10. These results show that the amount of evolutionary change associated with genetic drift is inversely related to the population size. When such a proportional deviation is small (*i.e.*, the population size is large), the *Hardy-Weinberg* Law predicts that gene frequencies and genotype ratios in the population reach an equilibrium and remain at that point from generation to generation. In other words, a population which meets the seven criteria will not evolve at all; or in a large population.

More interestingly, in any finite population one of the given alleles will ultimately dominate the population. This is exactly because of sampling error. Thus the pace of single allele fixation is slow in a large population. Since in the real world it is highly unlikely to meet any of those conditions (especially infinite population and no mutation are physically impossible conditions), a population naturally evolves, and single allele fixation is expected. Fixation in a large population requires so long a period (remember that sampling error decreases as an inverse function of the population size), that some other evolutionary factor may well prevent from it. On the other hand, in a case of small populations where *Hardy-Weinberg* Law loosens its grip, fixation brings a salient consequence; any genetic drift must cease at some realistic point in the population's evolution. If we apply a realistic condition; introducing mutation into the genepool model, fixation means that for any mutation,

its fate is either total dominance or extinction on the given locus in the population. This is shown in Figure A.1 (p. 218)[1].

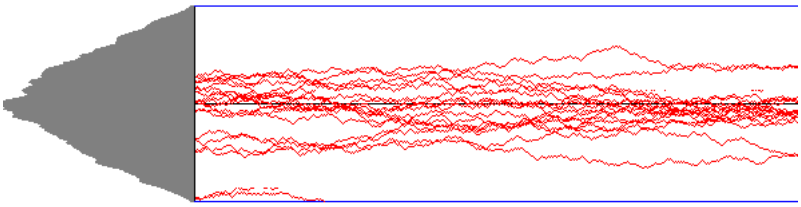

Figure A.1: 1D-Random Walk

This figure is a snapshot of a bounded random walk in one dimension. The red line starts from the left edge of the blue center line. The a random walk (a red line) proceeds horizontally for one unit per time step. When it reaches the right edge, it turns back to the left. Moving up or down is random (the size of movement in one step is also fixed). When it reaches either the top or bottom boundary, a new random walk starts from the center. The histograms to the left of the walk show the frequency of visited points on the dimension. The width of the boundaries corresponds to the size of population. Reaching the boundaries is equal to genetic fixation (reaching the top represents the given allele attaining complete dominance). If the width is infinitely wide (*i.e.*, infinitely large population), apparently no fixation is reached. A wide boundary means random walk will require large amount of time steps to reach the boundaries. Histograms reveal an interesting property; if the process continues, the shape of the histogram approximates to a Gaussian distribution (the gray, triangle figure in Figure A.1).

It is important to note, however, that random walk (and genetic drift) is not necessarily undirectional as often misunderstood. Suppose genetic drift changes the allele frequency from 0.5 to 0.6 in a particular population. Since genetic drift is random, one might expect in the next generation, the allele frequency may come back to the original. However, this is unlikely to be the case. Drift at a given generation is always around the previous generation's allele frequency. In other words, any given generation's allele frequency is affected only by the previous generation but not by more ancient generations. Thus it is fallacious to consider that allele frequencies tend to return to their ancestral frequencies. This is quite different from, say, the case of tossing coin. The frequency to have head or tail is always 0.5; and the current result of tossing coin has no effect on the probability to have a head in

[1]All figures shown this chapter are generated by using Dr. M. Burge's java applets on his homepage at Armstrong Atlantic State University (http://vision.armstrong.edu/burge/53.0.html, at this time of writing –May 2004)

next tossing. In the allele frequency case, once the allele frequency changes to 0.6, deviation will not be equally likely above and below 0.5. It is more likely to stay above 0.5. Thus with increases in generation numbers, it becomes more and more likely that the allele frequency will depart from the original deviation. Therefore, changes caused by genetic drift accumulate over the time.

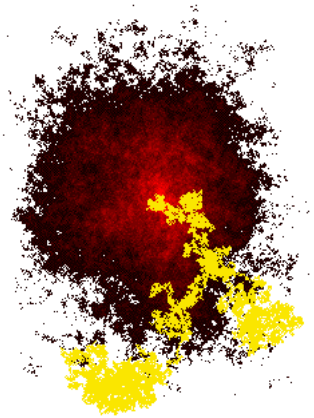

Figure A.2: 2D-Random Walk

Subsequently, genetic drift may potentially produce a striking evolutionary profile. Figure A.2 (p. 219) is a snapshot of a 2D-random walk simulation. Each time-step, a line, starting from the center of the graph, proceeds in four possible directions up, down, right, or left. When the line reaches the boundary a new line starts from the center. A crimson-colored cloud-like background shows previous lines and the yellow line is the current random walk. As a whole, random walks uniformly diffuse from the center. Individual walks, however, sometimes produce more 'determined' movements. A typical case of this is the ongoing random walk in the graph. Crucially, when comparing two different population sizes, say N=10 and 20, while changes can be seen in both populations, the degree of these changes is more pronounced in the smaller population. Crucially, in reality evolutionary processes are non-repetitive and temporally limited anyway, the window size of a random walk is necessarily small. Subsequently, a random but directed genetic drift may be observed in an evolutionary profile. This implies that it may consequently have a large impact on later generations.

Neutrality is often discussed with a consideration of fitness landscape. As increasing fitness corresponds to climbing a hill on a fitness landscape, neutral evolution is equal to walking on a level plane in the landscape. In other words, selection can cull individuals only when they climb up or down the hill; as long as they are walking on the level plane, they are not subject to selection.

If we trace individuals' movement on the plane, it might be quite directed as seen above. When such a movement goes extreme, it may reach an edge of the plane. If an individual moves across the edge outside the plane, obviously the walk will suddenly become not random. This itself is not particularly interesting. However, if we shift our attention to the population level from the individual level, an intriguing perspective comes up. While each family lineage may have a quasi-directed random walk, as the population, it is obvious that random walks as a whole may uniformly diffuse from the starting point; this is especially likely when the population size is large. Figure A.2 shows this aspect as well. As described above, when the diffusion reaches the edge of the plateau, culling by natural selection begins. Naturally, not all regions beyond the edges have the same inclinations; some may be downward, and others may be upward. Besides, the shape of diffusion may not be perfectly uniform, rather it is more likely to be skewed. In these cases, a small number of individuals who come across a certain edge of the plateau at a positive and steep inclination can be suddenly selected and become highly prolific. Subsequently, in later generations, the distribution of alleles' frequency can be quite different from that of the previous generations. This is called "the founder effect". Similar things can happen when a given environment rapidly changes or exaptation takes place. If environmental changes are radical, only a minor number of individuals survive to breed. Especially, in the consideration of neutrality, imagine the case that a rapid environmental change in some case corresponds to a sudden shrinkage of the plateau. Under the new environment, some of the plateau turns out to be no more flat; some parts may be lower than the original, and others may be higher. If a small number of individuals who are on the new higher positions become highly prolific and rest of the population are not able to survive to breed, then those individuals' alleles are highly likely to be very frequent in later generations. This, a very similar situation the founder effect, is called "the genetic bottleneck effect".

At the population level, individuals' random walks caused by neutral mutations increase the population's entropy; that is, genetic diversity of the genepool increases. Also this means that the distribution of (over the fitness landscape) of individuals increases. The population might begin a random walk, that is more and more neutral mutations accumulate in the population. At some point, however, such a random walk may have to come to the end. Any benign environments have a threshold at which point they start winnowing out further variants. Thus when the population on a flat land, they start to random walks. This is one of the key concepts in the recent development of the idea called "neutral networks" in computer science.

In sum, the following properties of genetic drift (random walk) can be noted.

No Direction When allele frequencies are averaged over generations, almost no prediction can be made from the initial allele frequencies.

Accumulation with Time The chances of any subpopulation deviating from the initial allele frequencies and the magnitude of that deviation increase each generation.

The Loss of Genetic Variability Fixation of one allele in a population is inevitable; either completely loss of the allele complete or domination of the population.

A.2 Evolutionary Spandrels

The other important process which provides genetic diversity is spandrels, introduced by Gould & Lewontin (1979). The word 'spandrel' apparently comes from an architectural term, the triangular space 'left over' between a rounded arch and the rectangular frame of wall and ceiling. By citing this, Gould describes that some of an organism's traits (indeed, he thinks of a non-trivial amount of organism's traits) are non-adaptive. It is a non-adaptive byproduct that subsequently appears as a consequence of the evolution of other adaptive traits. Gould criticizes that a number of researchers are often *ultra*-panselectionists; they often attempt to find adaptive reasons for every trait found in organisms. Gould's spandrel theory is an antithesis of such an extreme, but an often overlooked view.

Some spandrels would be immediately 'meaningful' for natural selection so that some selection process on the spandrel may quickly take place. However, some other spandrels (if not most) may be neutral so that it is evolutionary 'invisible'. Among such invisible spandrels, some would be assigned a new adaptive value. This change would be caused by autonomous environmental change or evolution of other traits. In any case, when a new environment appears, some spandrels may well become adaptive and as a consequence a new evolutionary process on the spandrel would take place. Thus this is also a case of exaptation. Apparently, in BNC, environmental changes are caused by the niche construction process.

Regarding language evolution, the spandrel theory may provide a more plausible scenario than the neutral theory. The human brain is structurally so complex, it would be of no surprise if plenty of spandrels (both structurally and functionally) existed. This is the reason that Gould explicitly expresses that language is a case of exaptation of spandrels in our brain. He considers that while our brain has become computationally powerful (in the evolutionary perspective), this is not because language requires such a power so that it becomes a selective pressure, but it has coopted such a power; evolutionary reasons of our big brain are found in different places.

Often this exaptation theory based on spandrels in the brain is confused with a so-called "big-bang" theory of language evolution. Nativists especially are eager to maintain their view of dysfunctionality of language, generally applaud this big-bang theory of language emergence since they think that this view is essentially the same as their 'non-adaptive' theory of language.

The logic behind their assumption seems to be fairly consistent; if language is non-functional, it would also not be evolutionary adaptive. Then some non-adaptive theory of language evolution (or language origin) will be required. Gould's theory seems to fit beautifully. However, this 'adaptive or not" type of linguistic argument which is often brought by linguists does not conform to Gould's spandrel theory of language evolution; languages can be still adaptive in evolutionary sense and exaptation would have been supported by various adaptive reasons. What Gould is highly skeptical of is the attitude to provide extensive evolutionary reasons *linguistically* toward both biological and psychological foundations of linguistic ability. In other words, he criticizes the attitude that applies the idea of domain-specificity of linguistic ability to the domain of evolutionary account –*X* is a domain-specific ability so that it has *domain-specific* evolutionary foundations[2]. When the adaptive view was introduced (like in generative linguistics), this caution was overshadowed and actually adopted in the opposite way.

Fundamentally, the spandrel scenario does not logically deny a gradual, accumulative theory of language evolution like we have considered. Gould also admits that language has been one of the major selective pressures for the brain to be more computationally powerful. Although this may sound somewhat contradictory, it is completely logical, or even more plausible; initially language used exaptation so that some 'spandrels' in the brain were coopted. Then, later on normal evolution acted on on such spandrels (of course, now it is no more non-adaptive). In short, this view of language evolution can be viewed as an "exaptation first, evolution next" type of evolutionary theory.

[2]This idea is found in debates between Pinker and Gould (*e.g.*, Gould 1997, Kaland *et al.* 1997).

References

Ancel L W (1999). A Quantitative Model of the Simpson-Baldwin Effect, *Journal of Theoretical Biology* **196**, 197–209.

Ancel L W (2000). Undermining the Baldwin Expediting Effect: Does Phenotypic Plasticity Accelerate Evolution?, *Theoretical Population Biology* **58**, 307–319.

Arita T (2000). *Artificial Life: a Constructive Approach to the Origin/Evolution of Life, Society, and Language*, Science Press, Tokyo. (in Japanese).

Arthur W B (1994). *Increasing Returns and Path Dependence in the Economy*, University of Michigan Press, Ann Arbor, MI.

Avital E & Jablonka E (2000). *Animal Traditions: Behavioural Inheritance in Evolution*, Cambridge University Press.

Baker C L (1979). Syntactic Theory and the Projection Problem, *Linguistic Inquiry* **10**, 533–581.

Baldwin J M (1895). Consciousness and Evolution, *Science* **2**, 219–223.

Baldwin J M (1896*a*). Consciousness and Evolution, *Psychological Review* **3**, 300–308.

Baldwin J M (1896*b*). Heredity and Instinct (I), *Science* **3**, 438–441.

Baldwin J M (1896*c*). Heredity and Instinct (II), *Science* **3**, 558–561.

Baldwin J M (1896*d*). A New Factor in Evolution, *The American Naturalist* **30**, 441–451, 536–553.

Baldwin J M (1896*e*). On Criticisms of Organic Selection, *Science* **4**, 727.

Baldwin J M (1902). *Development and Evolution*, Macmillan, New York, NY.

Bates E A, Elman J L, Johnson M H, Karmiloff-Smith A, Parisi D & Plunkett K (1998). Innateness and Emergentism, *in* W Bechtel & G Graham, eds, *A companion to cognitive science*, Blackwell, Oxford, England, pp. 590–601. **URL:** *http://crl.ucsd.edu/ bates/papers/pdf/bates-elman-johnson-karmiloff-s*

Belew M L (1990). Evolution, Learning, and Culture: Computational Metaphors for Adaptive Algorithms, *Complex Systems* **4**, 11–49.

Best M L (1999). How culture can guide evolution, *Adaptive Behavior* **7**(3/4), 289–306.

Bickerton D (1990). *Language and Species*, University of Chicago Press, Chicago.

Block N J (1981). Readings in the Philosophy of Psychology, the language and though series edn, Vol. 2, Harvard University Press, Cambridge, MA.

Boyd R & Richerson P H (1985). *Culture and the Evolutionary Process*, University of Chicago Press.

Briscoe E J (1997). Co-Evolution of Language and of the Language Acquisition Device, *in* P R Cohen & W Wahlster, eds, *Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, Association for Computational Linguistics, Somerset, New Jersey, pp. 418–427.

Briscoe E J (1998), 'Language as a complex adaptive system: co-evolution of language and of the language acquisition device'.
URL: *http://citeseer.nj.nec.com/briscoe98language.html*

Briscoe E J (1999*a*). Bayesian Learning of (Stochastic) Grammars, *in* J Cussens, ed., *Proceedings of the 1st Workshop on Learning Language in Logic*, Bled, Slovenia, p. 6. Abstract.

Briscoe E J (1999*b*). Grammatical Acquisition: Co-Evolution of Language and the Language Acquisition Device, *in Evolutionary Perspectives on Diachronic Syntax: Proceedings of the Diachronic Generative Syntax*, Vol. 5, Oxford University Press, Oxford, England.

Briscoe E J (2000*a*). Evolutionary Perspectives on Diachronic Syntax, *in* S Pintzuk, G Tsoulas & A Warner, eds, *Diachronic Syntax: Models and Mechanisms*, Oxford University Press, Oxford, England, pp. 75–108.

Briscoe E J (2000*b*). Grammatical Acquisition: Inductive Bias and Coevolution of Language and the Language Acquisition Device, *Language* **76**(2), 245–296.
URL: *http://citeseer.nj.nec.com/briscoe00grammatical.html*

Briscoe E J (2002*a*). Coevolution of the Language Faculty and Language(s) with Decorrelated Encodings, *in* M Tallerman, ed., *Proceedings of 4th International Conference on Evolution and Language*, Oxford University Press, Oxford, England.

Briscoe E J (2002*b*). Grammatical Acquisition and Linguistic Selection, *in* E J Briscoe, ed., *Language Acquisition and Linguistic Evolution: Formal and Computational Approaches*, Cambridge University Press, Cambridge, England.

Briscoe E J (2003). Grammatical Assimilation, *in* M Christiansen & S Kirby, eds, *Language Evolution: the States of the Art*, Oxford University Press, Oxford, England.

Brodie III E D (2000). Why Evolutionary Genetics Does not Always Add up, *in* J B Wolf, E D Brodie III & M J Wade, eds, *Epistasis and the Evolutionary Process*, Oxford University Press, Oxford, England, chapter 1, pp. 3–19.

Cangelosi A & Harnad S (2002). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories, *Evolution of Communication* **4**(1), 117–142.
URL: *http://www.tech.plym.ac.uk/soc/research/neural/staff/acangelosi/papers/cangelosi evocom.ps.zip*

Chomsky N (1965). *Aspects of the Theory of Syntax*, The MIT Press, Cambridge, MA.

Chomsky N (1972). *Language and Mind*, Harcout, Brace and Jovanovich, New York, NY.

Chomsky N (1975). *Reflections on Language*, Pantheon Press.

Chomsky N (1981). *Lectures on Government and Binding*, Foris, Dordrecht, The Netherlands.

Chomsky N (1982*a*). Discussion of Putnam's Comments, *in* M Piattelli-Palmarini, ed., *Language and Learning: the Debate Between Jean Piaget and Noam Chomsky*, Harvard University Press, Cambridge, MA.

Chomsky N (1982*b*). *The Generative Enterprise: A Discussion with Riny Huybregts and Henk van Riemsdijk*, Foris, Dordrecht, The Netherlands.

Chomsky N (1988). *Language and Problems of Knowledge*, The MIT Press, Cambridge, MA.

Chomsky N (1991). Linguistics and Cognitive Science, *in* A Kasher, ed., *The Chomskyan Turn: Generative Linguistics, Philosophy, Mathematics, and Psychology*, Blackwell, London, England.

Clark R (1994). Finitude, Boundedness and Complexity: Learnability and the Study of First Language Acquisition, *in* B Lust, M Suner & G Hermon, eds, *Syntactic Theory and First Language Acquisition: Crosslinguistic Perspectives*, Vol. II, Lawrence Erlbaum, Mahwah, NJ, pp. 473–789.

Clausen J, Keck D D & Hiesey W M (1948). 'Experimental studies on the nature of species. III. Environmental responses of climatic races of Achillea.', Carnegie Institution of Washington, publication 581, Washington, D.C.

Clausen J, Keck D D & Hiesey W M (1958). 'Experimental studies on the nature of species. IV. Genetic structure of ecological races.', Carnegie Institution of Washington Publication 615, Washington, D.C.

Corning P A (1983). *The Synergism Hypothesis: a Theory of Progressive Evolution*, McGraw-Hill, New York, NY.

Corning P A (1998). The Synergism Hypothesis: the Concept of Synergy and It's Role in the Evolution of Complex Systems, *Journal of Social and Evolutionary Systems* **21**(2), 133–172.

Curtiss S (1977). *Genie: A Psycholinguistic Study of a Modern Day "Wild Child"*, Academic Press, New York, NY.

Darwin C (1881). *The Formation of Vegetable Mould, Through the Action of Worms, with Observations on Their Habits*, Murray, London, England.

Dawkins R (1976). *The Selfish Gene*, Oxford University Press, Oxford, England.

Dawkins R (1982). *The Extended Phenotype*, Oxford University Press, Oxford, England.

Deacon T W (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*, W.W. Norton.

Deacon T W (2003). Multilevel Selection in a Complex Adaptive System: The Problem of Language Origins, *in* B H Weber & D J Depew, eds, *Evolution and Learning*, The MIT Press, Cambridge, MA, chapter 5, pp. 81–106.

Dennett D (1991). *Consciousness explained*, Little, Brown and Co., Boston, MA.

Dennett D (1995). *Darwin's Dangerous Idea*, Simon and Shuster, New York, NY.

Dennett D (2003). The Baldwin Effect: a Crane, not a Skyhook, *in* B Weber & D Depew, eds, *Evolution and Learning: the Baldwin Effect Reconsidered*, The MIT Press, Cambridge, MA.

Depew D (2000). The Baldwin Effect: an Archeology, *Cybernetics and Human Knowing* **7**(1), 7–20.

Dor D & Jablonka E (2000). From Cultural Selection to Genetic Selection: A Framework for the Evolution of Language, *Selection* **1**(1-3), 33–55.
URL: *http://www.akkrt.hu*

Dor D & Jablonka E (2001). How Language Changed the Genes: Toward an Explicit Account of the Evolution of Language, *in* J Trabant & S Ward, eds, *New Essays on the Origin of Language*, Vol. 133 of *Trends in Linguistics. Studies and Monographs*, Mouton de Gruyter, Berlin, chapter 9, pp. 151–175.

Downes S (2003). Baldwin Effect and the Explanatory Repertoire in Evolutionary Biology, *in* B Weber & D Depew, eds, *Evolution and Learning: the Baldwin Effect Reconsidered*, The MIT Press, Cambridge, MA.

Dryer M S (1989). Large Linguistic Areas and Language Sampling, *Studies in Language* **13**, 257–292.

Durham W (1991). *Coevolution: Genes, Culture and Human Diversity*, Stanford University Press.

Ellefson M R & Christiansen M H (2000*a*). The Evolution of Subjacency without Universal grammar: Evidence from Artificial Language Learning, *in* J L

Dessalles & L Ghadakpour, eds, *The Evolution of Language: 3rd International Conference*, Ecole Nationale Superieure des Telecommunications, Paris, France, pp. 98–101. Populations: Models and Algorithms, R. K. Belew and M. Mitchell (eds.), 1996, pp. 91–97, Reading, MA: Addison Wesley.

Ellefson M R & Christiansen M H (2000*b*). Subjacency Constraints without Universal Grammar: Evidence from Artificial Language Learning and Connectionist Modeling, *in The Proceedings of the 22nd Annual Conference of the Cognitive Science*, Lawrence Erlbaum, Mahwah, NJ, pp. 645–650.

Elman J L, Bates E A, Johnson M H, Karmiloff-Smith A, Parisi D & Plunkett K (1996). *Rethinking Innateness: a Connectionist Perspective on Development*, The MIT Press, Cambridge, MA.

Erdos P & Renyi A (1959). On Random Graphs, *Publicationes Mathematicae* **6**, 290–297.

Fisher J & Hinde R A (1949). The Opening of Milk Bottles by Birds, *British Birds* **42**, 347–357.

Fodor J (1983). *The Modularity of Mind*, The MIT Press, Cambridge, MA.

Fodor J (1989). Learning the Periphery, *in* R J Matthew & W Demopoulos, eds, *Learnability and Linguistic Theory*, Kluwer.

Godfrey-Smith P (2003). Between Baldwin Skepticism and Baldwin Boosterism, *in* B Weber & D Depew, eds, *Evolution and Learning: the Baldwin Effect Reconsidered*, The MIT Press, Cambridge, MA, chapter 3, pp. 53–67.

Goldshmidt R B (1938). *Physiological Genetics*, McGraw-Hill, NY.

Gottlieb G (1971). *Development of Species Identification in Birds: An Inquiry into the Prenatal Determinants of Perception*, University of Chicago Press, Chicago.

Gould S J (1991). Exaptation: a Crucial Tool for an Evolutionary Psychology, *Journal of Social Issues* **47**, 43–65.

Gould S J (1997). Darwinian Fundamentalism, *The New York Review of Books* **44**(10).

Gould S J & Eldredge N (1977). Punctuated equilibria: the tempo and mode of evolution reconsidered, *Paleobiology* **3**, 115–151.

Gould S J & Lewontin R (1979). The Spandrels of San Marco and the Panglossian Paradigm, *Proceedings of the Royal Society of London* (205), 281–288.

Gould S J & Vrba E S (1982). Exaptation: A missing term in the science of form, *Paleobiology* **8**(1), 4–15.

Gray R D (1992). Death of the Gene: Developmental Systems Strike Back, *in* P E Griffiths, ed., *Trees of Life*, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 165–209.

Greenberg J H (1963). Some Universals of Language with Special Reference to the Order of Meaningful Elements, *in* J H Greenberg, ed., *Universals of Language*, The MIT Press, Cambridge, MA, pp. 73–113.

Griffiths P E (2001). Evo-Devo Meets the Mind: Towards a Developmental Evolutionary Psychology, *in* R Brandon & R Sansom, eds, *Integrating Evolution and Development: A Collection on the Developmental Synthesis*, Cambridge Series in Philosophy of Biology, Cambridge Unversity Press, Campbridge, England.

Griffiths P E (2002). What is Innateness?, *The Monist* **85**(1), 70–85.

Griffiths P E & Gray R D (1994). Developmental Systems and Evolutionary Explanation, *Journal of Philosophy* **91**(6), 277–304.

Griffiths P E & Gray R D (2001). Darwinism and Developmental Systems, *in* S Oyama, P E Griffiths & R D Gray, eds, *Cycles of Contingency: Developmental Systems and Evolution*, The MIT Press, Cambridge, MA, chapter 16, pp. 195–218.

Grimshaw J (1981). Form, Function, and the Language Acquisition Device, *in* C L Baker & J J McCarthy, eds, *The Logical Problem of Language Acquisition*, The MIT Press, Cambridge, MA, pp. 165–182.

Harvey I (1993). The Puzzle of the Persistent Question Marks: A Case Study of Genetic Drift, *in* S Forrest, ed., *Proceedings of the Fifth International Conference on Genetic Algorithms*, Morgan Kaufman, San Mateo, CA, pp. 15–22. **URL:** *http://citeseer.nj.nec.com/harvey93puzzle.html*

Hinton G & Nowlan S (1987). How Learning Can Guide Evolution, *Complex Systems* **1**, 495–502. reprint in: Adaptive Individuals in Evolving Populations: Models and Algorithms, R. K. Belew and M. Mitchell (eds.), 1996, pp. 447–454, Reading, MA: Addison Wesley.

Holland J H (1975). *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI.

Hurford J (1990). Nativist and Functional Explanations in Language Acquisition, *in* I M Roca, ed., *Logical Issues in Language Acquisition*, Foris, Dordrecht, The Netherlands.

Jackendoff R (1983). *Semantics and Cognition*, The MIT Press, Cambridge, MA.

Jackendoff R (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*, Oxford University Press, Oxford.

Johnson M H & Morton J (1991). *Biology and Cognitive Development: The Case of Face Recognition*, Blackwell, Oxford, England.

Kaland H, Pinker S & Kalow W (1997). Evolutionary Psychology: An Exchange, *The New York Review of Books* **44**(15).

Karmiloff-Smith A, Plunkett K, Johnson M H, Elman J L & Bates E A (1998). What Does It Mean to Claim that Something Is 'Innate'? Response to Clark, Harris, Lightfoot and Samuels, *Mind and Language* **13**(4), 588–597.
URL: *http://crl.ucsd.edu/ elman/Papers/mind_and_language.pdf*

Kauffman S (1995). *At Home in the Universe: The Search for Laws of Complexity*, Oxford University Press, Oxford, England.

Kauffman S A (1989). Adaptation on Rugged Fitness Landscapes, *in* D L Stein, ed., *Lectures in the Sciences of Complexity*, Vol. I of *SFI Studies in the Sciences of Complexity*, Addison Wesley, Redwood City, CA, chapter 12, pp. 527–618.

Kimura M (1983). *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge, England.

Kirby S (2000). Syntax without Natural Selection: How compositionality emerges from vocabulary in a population of learners., *in* C Knight, ed., *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, Cambridge University Press, pp. 303–323.
URL: *http://www.cs.utexas.edu/users/hewett/research/airg/evol98.ps.gz*

Kirby S (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity, *IEEE Transactions on Evolutionary Computation* **5**(2), 102–110.
URL: *http://www.ling.ed.ac.uk/anonftp/pub/staff/kirby/skirbyieee.ps*

Kirby S (2002). Learning, bottlenecks and the evolution of recursive syntax, *in* E Briscoe, ed., *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, Cambridge University Press, Cambridge, England, pp. 173–203.

Kirby S & Hurford J (1997). Learning, Culture and Evolution in the Origin of Linguistic Constraints, *in* P Husbands & I Harvey, eds, *Proceedings of the Fourth European Conference on Artificial Life (ECAL-97)*, A Bradford Book, The MIT Press, Cambridge, MA, pp. 493–502.

Kirby S & Hurford J R (2002). The Emergence of Linguistic Structure: An overview of the Iterated Learning Model, *in* A Cangelosi & D Parisi, eds, *Simulating the Evolution of Language*, Springer Verlag, London, UK.

Laland K N, Odling-Smee F J & Feldman M W (1996). On the Evolutionary Consequence of Niche Construction, *Journal of Evolutionary Biology* **9**, 293–316.

Laland K N, Odling-Smee F J & Feldman M W (1999). Evolutionary Consequences of Niche Construction and Their Implications for Ecology, *in Proceedings of the National Academy of Sciences*, Vol. 96, pp. 10242–10247.

Laland K N, Odling-Smee F J & Feldman M W (2000). Niche construction, biological evolution and cultural change, *Behavioral and Brain Science* **23**, 131–175.

Laland K N, Odling-Smee F J & Feldman M W (2001*a*). Cultural Niche Construction and Human Evolution, *Journal of Evolutionary Biology* **14**, 22–33.

Laland K N, Odling-Smee F J & Feldman M W (2001*b*). Niche Construction, Ecological Inheritance, and Cycles of Contingency in Evolution, *in* S Oyama, P E Griffiths & R D Gray, eds, *Cycles of Contingency: Developmental Systems and Evolution*, The MIT Press, Cambridge, MA, chapter 10, pp. 117–126.

Lehrman D S (1953). A Critique of Konrad Lorenz's Theory of Instinctive Behavior, *Quarterly Review of Biology* **28**, 337–363.

Lenneberg E H (1967). *Biological Foundations of Language*, John Wiley and Sons.

Lewontin R (2000). *The Triple Helix: Gene,Organism, and Environment*, Harvard University Press, Cambridge, MA.

Lewontin R C (1974). The Analysis of Variance and the Analysis of Causes, *American Journal of Human Genetics* **26**, 400–411.

Lewontin R C (1983). Gene, Organism, and Environment, *in* D S Bendall, ed., *Evolution: From Molecules to Men*, Cambridge University Press, Cambridge, England, pp. 273–285.

Liberman A & Mattingly I (1989). A Specialization for Speech Perception, *Science* **243**, 489–496.

Marr D (1982). *Vision*, W. H. Freeman and Company, San Francisco, CA, USA.

Mayley G (1996*a*). The Evolutionary Cost of Learning, *in* P Maes, M J Mataric, J A Meyer, J Pollack & S W Wilson, eds, *From Animals to Animats 4: Proceeding of the Fourth International Conference on Simulation of Adaptive Behavior*, The MIT Press, Cambridge, MA.

Mayley G (1996*b*). Landscapes, Learning Costs and Genetic Assimilation, *Evolutionary Computation* **4**(3), 213–234. Special edition: Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect.

Mayley G (1997). Guiding or Hiding: Explorations into the Effects of Learning on the Rate of Evolution, *in* P Husbands & I Harvey, eds, *Proceedings of Fourth European Conference on Artificial Life (ECAL-97)*, A Bradford Book, The MIT Press, Cambridge, MA.

Maynard Smith J H (1982). *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, England.

Maynard Smith J H (1987). Natural Selection: When Learning Guides Evolution, *Nature* **329**, 761–762. Reprint in: Adaptive Individuals in Evolving Populations: Models and Algorithms, R. K. Belew and M. Mitchell (eds.), 1996, pp. 455–456, Reading, MA: Addison Wesley.

Maynard Smith J H & Haigh J (1974). The Hitch-Hiking Effect of a Favorable Gene, *Genet. Res.* **23**, 23–35.

Morgan T H (1926). *The Theory of the Gene*, Yale University Press, New Haven, CT.

Newmeyer F J (2000). On the reconstruction of 'Proto-world' word order, *in* C Knight, J R Hurford & M Studdert-Kennedy, eds, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, Cambridge University Press, Cambridge, England.

Odling-Smee F J (1988). Niche Constructing Phenotypes, *in* H Plotkin, ed., *The Role of Behavior in Evolution*, The MIT Press, Cambridge, MA.

Odling-Smee F J (1994). Niche Construction, Evolution and Culture, *in* T Ingold, ed., *Comapnion Encyclopedia of Anthropology*, Routledge, London, England.

Odling-Smee F J, Laland K N & Feldman M W (1996). Niche Construction, *The American Naturalist* **147**(4), 641–648.

Odling-Smee F J, Laland K N & Feldman M W (2003). *Niche Construction: The Neglected Process in Evolution*, number 37 *in* 'Monographs in Population Biology', Princeton University Press, Princeton, NJ.

Oyama S (1985). *The Ontogeny of Information: Developmental Systems and Evolution*, Cambridge University Press, Cambridge, England.

Oyama S, Griffiths P E & Gray R D (2001). What is Developmental Systems Theory?, *in* S Oyama, P E Griffiths & R D Gray, eds, *Cycles of Contingency: Developmental Systems and Evolution*, A Bradford Book, The MIT Press, Cambridge, MA, chapter 1, pp. 1–11.

Piatelli-Palmarini M (1989). Evolution, Selection, and Cognition: From "Learning" to Parameter Setting in Biology and the Study of Language, *Cognition* **31**, 1–44.

Pigliucci M (n.d.), 'Plasticity 101'.
URL: *http://fp.bio.utk.edu/pgl/handouts/plasticity_101.pdf*

Pinker S & Bloom P (1990). Natural language and natural selection, *Behavioral and Brain Sciences* **13**(4), 707–784.
URL: *http://www.bbsonline.org/documents/a/00/00/04/99/bbs00000499-00/bbs.pinker.html*

Rice S H (2000). The Evolution of Developmental Interactions, *in* J B Wolf, E D Brodie III & M J Wade, eds, *Epistasis and the Evolutionary Process*, Oxford University Press, Oxford, England, chapter 5, pp. 82–98.

Ross J (1967), Constraints on Variables in Syntax, PhD thesis, MIT, Cambridge, MA.

Samuels R (1998). What Brains won't Tell Us about the Mind: A Critique of the Neurobiological Argument Against Representational Nativism, **13**(4), 548–570.

Samuels R (2002). Nativism in Cognitive Science, **17**(3), 233–265.

Sasaki T & Tokoro M (1997). Adaptation toward Changing Environments: Why Darwinian in Nature?, *in Proceedings of Fourth European Conference on Artificial Life (ECAL-97)*, A Bradford Book, The MIT Press, Cambridge, MA, pp. 145–153.

Sasaki T & Tokoro M (1998). Adaptation under Changing Environments with Various Rates of Inheritance of Acquired Characters: Comparison between Darwinian and Lamarckian Evolution, *in Proceedings of 2nd Asia-Pacific Conference on Simulated Evolution And Learning (SEAL-98)*.

Sasaki T & Tokoro M (1999). Evolving Learnable Neural Networks under Changing Environments with Various Rates of Inheritance of Acquired Characters: Comparison between Darwinian and Lamarckian Evolution, *Artificial Life* **5**(3), 203–223.
URL: *http://citeseer.nj.nec.com/sasaki99evolving.html*

Schmalhausen I (1949). *Factors of Evolution: The theory of stabilizing selection*, University of Chicago Press, Chicago.

Shapiro J (2001). Genetic Algorithms in Machine Learning, *Lecture Notes in Computer Science* **2049**, 146–168.
URL: *http://citeseer.nj.nec.com/468697.html*

Simpson G (1953). The Baldwin Effect, *Evolution* **7**, 110–117. Reprinted in Adaptive Individuals in Evolving Population (1996).

Sober E (1992). Models of Cultural Evolution, *in* P E Griffiths, ed., *Trees of Life*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Steele E (1998). *Lamarck's Signature: How Retrogenes are Changing Darwin's Natural Selection Paradigm*, Perseus Books, Reading, MA.

Turkel W J (2002). The Learning Guided Evolution of Natural Language, *in* E J Briscoe, ed., *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, Cambridge University Press, Cambridge, England, chapter 8.

Turney P D (1996*a*). How to Shift Bias: Lessons from the Baldwin Effect, *Evolutionary Computation* **4**(3), 271–295.
URL: *http://citeseer.nj.nec.com/turney96how.html*

Turney P D (1996*b*). Myths and legends of the Baldwin Effect, *in Proceedings of the Workshop on Evolutionary Computing and Machine Learning at the 13th International Conference on Machin Learning (ICML96)*, Bari, Italy, pp. 135–142.

Turney P D, Whitley D & Anderson R (1996). Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect, *Evolutionary Computation* **4**(3), iv–viii.

Turney P D, Whitley L D & Anderson W R (1997). Introduction to the Special Issue: Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect, *Evolutionary Computation* **4**(3), ii–viii.

Waddington C H (1975). *The Evolution of an Evolutionist*, Edinburgh University Press, Edinburgh.

Wagner G, Booth G & Bagheri-Chainchian H (1997). A Population Genetic Theory of Canalization, *Evolution* **51**(2), 329–347.

Wexler K & Culicover P W (1980). *Formal Principles of Language Acquisition*, The MIT Press, Cambridge, MA.

Wiles J, Watson J, Tonkes B & Deacon T W (2002). Strange Loops in Learning and Evolution, *InterJournal Complex Systems* (572). Presented at the International Conference on Complex Systems (ICCS 2002).

Williams G C (1966). *Adaptation and Natural Selection*, Princeton University Press.

Wolf J B (2000). Indirect genetic effects and gene interactions, *in* J B Wolf, E D Brodie III & M J Wade, eds, *Epistasis and the Evolutionary Process*, Oxford University Press, Oxford, England, chapter 10, pp. 158–176.

Wolf J B, Brodie III E D, Cheverud J M, Moore A J & Wade M J (1998). Evolutionary Consequences of Indirect Genetic Effects, *Trends in Ecology and Evolution* **13**, 64–70.

Yamamoto Y, Sasaki T & Tokoro M (1999). Adaptability of Darwinian and Lamarckian Populations toward Unknown New World, *in Advances in Artificial Life: Proceedings of the 5th European Conference on Artificial Life (ECAL-99)*, Springer Verlag, Berlin, Germany, pp. 39–48.

Yamauchi H (1999), The Difficulty of Baldwinian Account of Language Evolution, Master's thesis, University of Edinburgh, Edinburgh, Scotland.

Yamauchi H (2001). The Difficulty of the Baldwinian Account of Linguistic Innateness, *in* J Kelemen & P Sosík, eds, *Advances in Artificial Life: Proceedings of the 6th European Conference on Artificial Life (ECAL-00)*, Lectures Notes in Computer Science, Springer Verlag, Berlin, Germany, pp. 391–400.
URL: *http://link.springer.de/link/service/series/0558/papers/2159/21590391.pdf*