Chapter 1

Strange loops in learning and evolution

Janet Wiles, James Watson, Bradley Tonkes
The University of Queensland
{j.wiles, jwatson, btonkes}@itee.uq.edu.au

Terrence Deacon
Harvard University
twdeacon@aol.com

Abstract

Scientific theories typically make sense of phenomena at a given level of explanation. Occasionally, phenomena that seem to belong to one level unexpectedly influence an entirely different one. These interactions are strange loops. In evolution and learning, one such strange loop, the Baldwin effect, was proposed over a century ago, and has been studied computationally for the past 15 years. In this study, we use computational techniques to explore a second strange loop, which we call genetic redistribution.

Deacon has recently proposed that complexes of genes can be integrated into functional groups as a result of environmental changes that mask and unmask selection pressures. For example, many animals endogenously synthesize ascorbic acid (vitamin C), but anthropoid primates have only a non-functional version of the crucial gene for this pathway. It is hypothesized that the loss of functionality occurred in the evolutionary past when a diet rich in vitamin C masked the effect of the gene, and its loss effectively trapped the animals in a fruit-eating lifestyle. As a result, the complex of abilities that support this lifestyle were evolutionarily bound together, forming a multilocus complex.

In this study we use evolutionary computation simulations to explore the thesis that masking and unmasking can transfer dependence from one set of genes to many sets, and thereby integrate the whole complex of genes. We use the vitamin C example, but emphasize that in this study we investigate the computational

consistency of the thesis, rather than its biological credentials. We used a framework based on Hinton and Nowlan's 1987 simulation of the Baldwin effect. Additional gene complexes and an environmental parameter were added to their basic model, and the fitness function extended. The simulation clearly demonstrates the computational consistency of the genetic redistribution effect, showing an initial advantage of endogenously synthesized vitamin C, followed by transfer of the benefit to the complex of genes that together allow the acquisition of vitamin C from the environment.

As is well-known in the modelling community, the Baldwin effect only occurs in simulations when the population of agents is "poised on the brink" of discovering the genetically specified solution. Similarly, the redistribution effect occurs in simulation under specific initial conditions: too little vitamin C in the environment, and its synthesis it is never fully masked; too much vitamin C, and the abilities required to acquire it are not tightly integrated. Both the Baldwin and the redistribution effects can be considered strange loops in the interactions between genotype and phenotype in evolution.

The Baldwin effect has been hypothesized as a potential mechanism for developing language-specific adaptations like innate Universal Grammar and other highly modular capacities. Our simulations support Deacon's argument that the process is likely the inverse, and that the extensive neural and other anatomical consequences would not be in the form of specific innate adaptations. Instead, the power of symbolic communication as a masking agent should unmask selection on an extensive and highly distributed constellation of capacities that would collectively come under selection for their fractional contributions to the acquisition and use of language.

1.1. Introduction: From specific genes to multilocus complexes

The evolution of vitamin C dependency in humans has an interesting evolutionary history. Many animals have the ability to endogenously synthesize ascorbic acid (vitamin C). The crucial gene in this synthetic pathway is for an enzyme (LGO) that catalyses the last stage of synthesis of vitamin C. Anthropoid primates (monkeys, apes, and humans) don't synthesize their own vitamin C, and the question that arises is why not, when it seems so useful. It turns out that primates including humans do indeed have the remains of such a gene that is no longer expressed and has accumulated irreparable mutational damage (it is now a "pseudogene" and was identified by Japanese researchers in 1994 using a probe gene from rat). The loss appears to have happened about 43 million years ago, which is the time that some primates became diurnal, with consequent changes in lifestyle, including a diet containing significant amounts of fruit. A convenient explanation (or just-so-story) is that the increased fruit in the animals' diets provided ample regularly available vitamin C, reducing the selection pressure to maintain the function of the vitamin C producing gene. Once the gene was corrupted in this lineage, the animals were effectively addicted to fruit, and trapped in the fruit-eating lifestyle: obligate frugivores. At this point, all the abilities that had incidentally supported the

acquisition of vitamin C through diet were then placed under a much stronger selection pressure, causing many to evolve in such a way as to better insure the ubiquitous availability of this now essential nutrient.

In recent work, Deacon (2001) hypothesized that this kind of secondarily distributed selection pressure is a major force in evolution, integrating what were initially a diverse set of potentially unrelated skills (e.g. colour vision, tooth structure, taste preferences, in this case) into a metastable multilocus multiphenotype adaptive complex.

Our goal in this paper is not to examine the vitamin C theory per se, but rather to investigate the internal computational consistency of the "genetic redistribution" argument. We begin by reviewing the relationship between evolution and learning, in particular the Baldwin effect and the computational studies that have been used to demonstrate it, and explore the conditions under which it occurs. We then extend those simulations to study the genetic redistribution effect.

1.2. Strange loop #1: The Baldwin effect

The field of complex systems studies the relationships between different levels of description of systems, and tries to characterise how emergent properties at one level arise from their component parts. Conventional explanations don't usually have causality encompassing very different levels of a multilevel system.

In evolutionary theory, conventional accounts explain how genotypes give rise to phenotypes. But changes in phenotype due to use or disuse are not inherited. That is, the acquired behaviours of an individual are not passed on to its offspring. The muscles of the blacksmith are not inherited, only the potential to develop them. Similarly, the broken leg of the skier is not inherited by their children, although the potential for broken legs is inherited.

When phenomena seem to unexpectedly influence an entirely different level, Cohen and Stewart (1995) call it a "strange loop". They describe Waddington's "genetic assimilation" as one such effect in which phenotypic-level behaviour influences genetic specification. In this effect, behaviours that were learned or acquired in the remote past, become genetically fixed - not in the subsequent generation - but in the remote descendants. The process sounds Lamarkian, but a Darwinian explanation exists. Over evolutionary time, natural selection acts on phenotypes and under some conditions will act to increase variance (or plasticity – which in this context includes both physical and behavioural plasticity and learning). Under other circumstances natural selection will act to decrease variance:

1. *Increased variance or plasticity:* If a population is not completely optimized for its current environment, behavioural variance may initially increase the range of behaviours open to individuals. Behaviours that are initially at the extreme range of a population's phenotypic plasticity may provide disproportionate benefit to those who can discover and utilise them during their lifetime. Such plasticity is then selected for, until the entire population is comprised of individuals with such plasticity.

2. *Decreased variance or plasticity:* At this stage, selection will favour individuals that are better optimized to their environment with less variance. Small mutations that provide a head start for the phenotype towards good solutions will be favoured. The head start may be accomplished in a multitude of ways – they all come under the optimization term of "bias". Over time, biases may become so strong that virtually no environmental stimulus is needed, and the behaviour can be considered genetically specified. The mechanism that underlies genetic assimilation is thus differential selection of individuals who either need less learning or learn faster.

The idea that plasticity or variance can play a computational role in evolutionary search using purely Darwinian mechanisms was proposed by Baldwin (1896) and Morgan (1896). Their original motivation was to understand how phenotypic plasticity or learning could guide evolution (without resorting to Lamarkianism). This interaction between learning and evolution has come to be known as the Baldwin effect.

From an evolutionary search perspective, phenotypic plasticity is credited with the ability to explore phenotypic space considerably faster and cheaper than genetic search allows. Genetic assimilation provides a Darwinian mechanism that enables the outcomes of phenotypic search to be incorporated into the genome over evolutionary timescales.

A computational model of the Baldwin effect was first demonstrated by Hinton and Nowlan (1987). They designed a simple simulation in which learning could guide an evolutionary algorithm to solve a computational version of a needle-in-haystack task. A population of agents were initiated with random genomes that included alleles for both innate and plastic behaviour. Agents were selected for their ability to find a single high fitness phenotype during their "lifetimes". Without phenotypic plasticity, the search task is exponential in the number of genes. Populations that are large enough for one or a few individuals to solve the task genetically lose the solution through crossover before it can become genetically fixed. With plasticity, populations reliably find the high fitness phenotype and genetically assimilate it.

Since Hinton and Nowlan's original simulation, a generation of computational modellers have explored multiple facets of the effect (Harvey, 1993; French and Messinger, 1994; Turney, Whitley & Anderson 1996; Mayley, 1996). Complex interactions occur between the parameters in the simulations. Of particular importance to this study, is that finding and assimilating a solution to an evolutionary search task only occurs under quite stringent circumstances:

- In Hinton and Nowlan's original model, the population size times the space that an agent can search in its lifetime was approximately equal to the size of the entire phenotypic search space. This balance ensured that in the initial random population, at most a few individuals would find the needle phenotype. Thus, the population was poised on the brink of discovering the solution.
- Populations rarely eliminated learning completely. The fitness function used by Hinton and Nowlan has very little selection pressure for the complete assimilation of the needle-in-a-haystack task. Hitch-hiking genes from the initial

founders and genetic drift result in homozygous alleles within relatively few tens of generations and residual learning remains (Harvey, 1993).

- The selection algorithm (e.g., fitness proportional or tournament selection) makes a difference to the amount of residual learning and the speed of assimilation (Wiles, Schultz, Bolland, Tonkes & Hallinan, 2001).
- The Baldwin effect has been demonstrated with a variety of models, including neural networks, both single layer and multi-layer. The rate of environmental change and the complexity of the landscape affects the benefits of learning in a rugged landscape (Watson, Geard and Wiles, submitted).

The interactions between parameters in the Baldwin effect can be summarised in terms of two components, the benefits and costs of learning (Mayley, 1996). In the early stages of a simulation, phenotypic plasticity is a cost effective way to search the local fitness landscape compared to genetic search. At later stages, the costs of learning outweigh the benefits, when all individuals can learn. However, residual learning is only eliminated when the learning costs remain high.

In summary, the strange loop from phene to gene can be explained in terms of changing frequencies of alleles over a population of learners. Learning need not always produce one or both stages of the Baldwin effect, neither increasing the speed of finding a solution, nor resulting in genetic assimilation. Both stages of the Baldwin effect only occur when learning changes the selection of parents for the subsequent generation, first as an advantage, and then as a cost.

The conclusion we draw from these studies is that the rise and fall of learning is a highly specific transient dynamic in the strange loop that links evolution and learning, not a ubiquitous feature of phenotypic plasticity.

1.3. Strange loop #2: Genetic redistribution

Viewing the Baldwin effect as a strange loop from phenotype to genotype invites the possibility that other strange loops may occur. Deacon has recently proposed that complexes of genes can be integrated into functional groups as a result of environmental changes that mask and unmask selection pressures. The vitamin C story outlined above is an example of such a dynamic.

When endogenous synthesis of vitamin C is masked by its presence in the diet, the distributed selection pressure can be viewed as a “reverse Baldwin effect” in that abilities specified directly in the genome may become masked by both internal and external sources, including flexible behavioral abilities, and over time their genetic specification is lost (a complexity catastrophe in Kauffman’s terms). When this occurs, the individual becomes dependent on the external source or whatever else has provided the masking effect, and any phenotypic capacities that support this masking (e.g. by providing an externally redundant nutrient) become increasingly elaborated and integrated through positive selection pressure. Deacon describes these two effects as “masking” and “unmasking” of selection.

In this study we use evolutionary computation simulations to explore the thesis that masking and unmasking can transfer dependence from one set of genes to many

sets, and thereby integrate the whole complex of genes. We use the vitamin C example, but emphasize that in this study we investigate the computational consistency of the thesis and the conditions under which it occurs, not its biological credentials. The first goal of the simulation was to demonstrate the computational consistency of the genetic redistribution effect, showing an initial advantage of endogenously synthesized vitamin C, followed by transfer of the benefit to the complex of genes that together allow the acquisition of vitamin C from the environment. The second aim was to explore the conditions under which the effect occurred.

1.3.1. Details of the Model Design

The design is an evolutionary computation (EC) model of genetic change over generations. The architecture is based on Hinton and Nowlan's (1987) simulation. Additional gene complexes and an environmental parameter were added to their basic model, and the fitness function extended. The elegance of the architecture is that it uses their framework but shows a different effect by changing the environmental context.

The simulation was designed to explore how masking and unmasking by environmental factors can transfer dependence from one gene to many and thereby integrate whole complexes of genes. Hinton and Nowlan used a single environment which was constant throughout the simulation. In the current study, the environment is modelled as a parameter that specifies the maximum amount of vitamin C available in an agent's environment. The environment has three stages, starting at a very low base, then rising to more than required for daily use, then dropping to half the daily limit.

To benefit from the environment, the agent must have the full set of abilities to acquire vitamin C from its food. If even one of those abilities was missing, then no exogenous vitamin C was acquired. The total vitamin C for an agent is calculated as the sum of the endogenous and exogenous sources, up to a maximum value. Any vitamin C beyond the maximum is discarded.

Following Hinton and Nowlan, in the current study each ability was modelled using a chromosome of 20 genes per ability and the fitness contribution was based on a needle-in-a-haystack task. To distinguish between the genotype and phenotype, the expressed value of a gene will be referred to as a phene. Each gene had three alleles, correct (1), incorrect (0) or learnable (?). The correct and incorrect alleles were fixed, and could not change during an agent's lifetime. The learnable alleles were reset randomly to correct or incorrect phenes each day of an agent's life. Each phene thus had two "alleles", correct or incorrect. The agent added a fitness contribution for an ability only when all its phenes were correct (no distinction is made between whether phenes were innately correct, or correct through guessing). The fitness function reflected a penalty for the number of guesses taken to find the correct phenotype.

The analogy in the vitamin C scenario is that setting all alleles to '1' corresponds to the ability (A_0) to endogenously synthesize vitamin C. The additional abilities needed to acquire vitamin C from the environment were modelled as k additional

gene complexes (A_1, A_2, … A_k). Each additional ability was also modelled as the coordinated action of 20 genes (resulting in N=20(1+k) genes in the chromosome). Abilities A_1-A_k were initially independent and individually advantageous (e.g., colour vision, tooth structure, taste preference). Together they enabled the agent to acquire and utilise vitamin C from its environment.

The fitness function, F, includes terms for the total vitamin C and the sum of the additional abilities, $F = 1+(N-1)*(c+a)$ where c is total vitamin C from both endogenous and environmentally derived sources, with excess vitamin C discarded, and a is the total incidental fitness from the other abilities (see Box 1 for details of the fitness function).

Box 1. Fitness function

$$F = 1 + (N-1)(c + a)$$

where

N is the genome length of A_0

C is the total vitamin C (endogenous and acquired), given by $c = \text{Min}[c_0 + E*c_1, 1.0]$

E is the amount of vitamin C available in the environment

c_0 is the amount of endogenous vitamin C, and c_1 is the ability to acquire vitamin C, given by

$c_0 = \text{Max}[0, (M-2^x)/M]$ if A_0 has no 0s, and x is the number of ?s in A_0
0 otherwise

$c_1 = \text{Min}[r_1, r_2, r_3]$

M is the number of guesses

a is the total contribution of other abilities, $a = \lambda \sum r_i *$

λ is the abilities constant and scales the relative contributions of vitamin C and the other abilities

r_i* is the fitness contribution from ability Ai, given by

$r_i = \text{Max}[0, (M-2^y)/M]$ if A_i has no 0s, and y is the number of ?s in A_i
0 otherwise

$r_i* = \text{Min}[r_i, 0.75]$

Simulation parameters for this example:

N=20, λ =30/(N-1), population size = 1000, M=1000,
tournament selection with mutation and single-point crossover
Distribution of alleles (initial and mutation): 35% 0s, 25% 1s, 40% ?s. This distribution makes the learning task slightly harder than Hinton and Nowlan's.

1.3.2 Results

The time course of the simulations demonstrates the initial advantage of the endogenously synthesized vitamin C, followed by a transfer of the ability to the complex of genes that mask the effect. The simulation has three obvious stages:

1. *Pre-history (0-700 generations):* all 4 abilities are *independently* found.
2. *Masking (700-1200):* ample vitamin C in the environment results in rapid loss of ability A0.
3. *Unmasking (1200-2000):* vitamin C in the environment is reduced, resulting in a *co-ordinated* increase in A1-A3.

Stage 1 is the pre-history of the effect, in which the population is evolving all its abilities (see Figure 1a-f, generations 0-700). The ability to synthesize vitamin C, A_0, is the first ability acquired in this trial (Figure 1c), followed by A_1 to A_3 (Figure 1d-f). In other trials they are found in random order. The important point to note in this stage is that each ability is found independently, and there is no co-ordination in the timing of their fitness increases. All abilities have a small contribution to the total fitness, which is why they evolve in the first stage. When all three abilities A_1-A_3 are found, the population has the potential to utilise vitamin C in the environment, but in this first stage there is virtually no environmental source, and hence no additional benefit beyond the individual contributions to fitness.

Stage 2 is the masking stage and begins when the environment changes to include significant amounts of vitamin C (see Figure 1a-f, generations 700-1200). During this stage, the dietary vitamin C masks the contribution of endogenous vitamin C. Random mutations gradually accumulate in the genes of A_0, and the ability to synthesize vitamin C is lost. At this point, the population is addicted to exogenous vitamin C. Abilities, A_1-A_3, which initially evolved independently of one another, are now bound together. At this stage, their co-dependence is not directly visible from the graphs, but any further evolutionary change must take into account that their combined effect is essential in acquiring vitamin C. However, the excess vitamin C in the environment enables them to be less than optimal in their acquisition of it. In Figure 1c, at 700 generations, a slight drop in fitness is seen. This average reflects loss of robustness to mutation, with 75% of the population at maximum fitness and 25% with mutations.

Stage 3 is the unmasking stage and begins when the conditions change again, and the amount of vitamin C in the environment is reduced to half the maximum benefit (see Figure 1a-f, generations 1200-2000). There is now selection pressure for abilities A_1-A_3 to optimize their extraction of vitamin C from the environment, and a coordinated increase in their fitness is seen (shown as a coordinated increase in fitness over generations 1200-1250 in Figure 1d-f). The chance of an agent rediscovering the correct genes for A_0 is significantly less than at the start of the simulation due to convergence of the population.

To test the robustness of the simulation, we ran 100 trials using the parameters described above with tournament selection, and 100 trials using fitness proportional selection. In all trials using either selection method, if all four abilities had been discovered by the onset of the masking phase, masking was completely effective in rapidly reducing the fitness of A_0 to zero. Both selection methods had some trials that also blocked the rediscovery of A_0, showing the classic genetic redistribution effect illustrated in Figure 1.

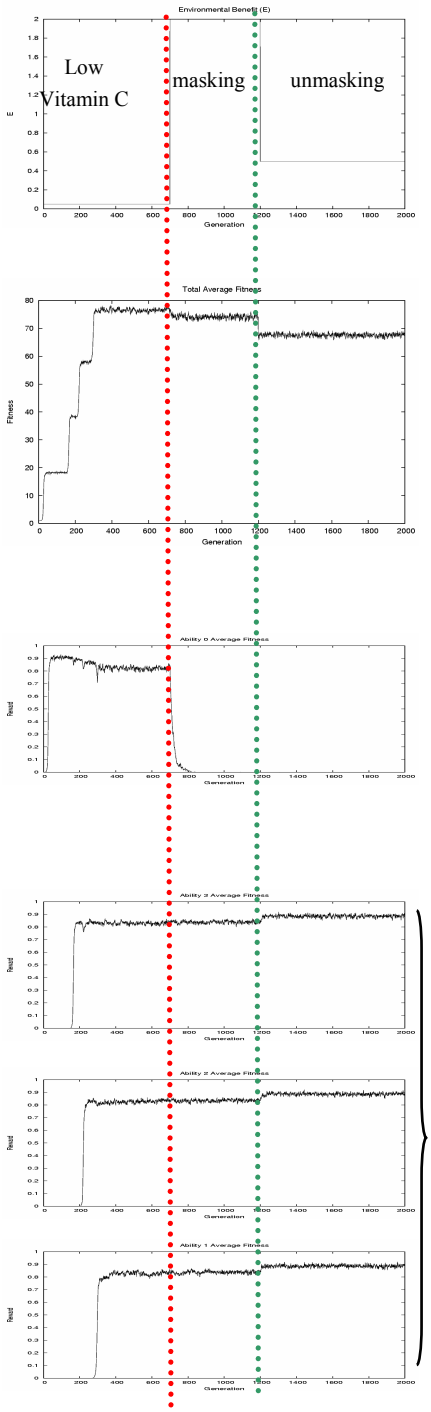

Figure 1a. Environmental benefit. For 700 generations, there is very little vitamin C in the environment. Then the amount rises to 200% of the usable amount. At 1200 generations it drops back to 50% of the usable amount.

Figure 1b. Total average fitness, *F*. The population starts with random alleles in the agent's genomes, and very low fitness. During stage 1, all four abilities are found. At 1200 generations, the fitness drops markedly, reflecting the reduction of vitamin C in the environment and a reduction of fitness over the population as a whole.

Figure 1c. Ability A_0 average fitness. A_0 is rapidly discovered and spreads throughout the population. At generation 700, A_0 fitness plummets, due to a lack of selection pressure to maintain the ability.

Figure 1d-f. Abilities A_1-A_3 average fitnesses. Abilities A_1-A_3 are discovered in the first 700 generations, but their fitness increases are uncorrelated with one another. At generation 1200, they show a correlated rise in fitness, optimising their extraction of vitamin C from the environment.

However, the simulations showed much other variation as well. In many trials all four abilities were not found before generation 1200, and hence there was no effect to observe, either because there was no A_0 to mask, or one of the other abilities was not present so no benefit could be gained from the environmental source of vitamin C. From previous experience with the Baldwin effect, it seems likely that hitch-hiking and rapid convergence of the population often prevent all four abilities being found in fitness proportional selection, and that convergence due to prolonged periods of drift and insufficient selection pressure may prevent all abilities being found within the given timeframe in tournament selection.

Of the trials that found all four abilities before the masking stage finished, the method of selection did show a difference in the proportion of times that vitamin C was blocked from re-evolving. Using tournament selection 30/100 trials found all four abilities within 1200 generations, and of those trials, 43% (13/30) blocked the return of A_0. Using fitness proportional selection, only 5/100 trials found all four abilities within 1200 generations, but all of those trials blocked the return of A_0.

1.4. Discussion and conclusions

The simulations clearly demonstrate the effects of masking and unmasking *in silico* and the consistency of the arguments for genetic redistribution. In particular, the two main aspects of the simulation were robust to a variety of parameters:

- Endogenous synthesis of vitamin C (A_0) was consistently lost due to masking by other abilities and an alternative source in the environment. Given sufficient vitamin C in the environment, loss of A_0 was complete in all cases. If the source of vitamin C in the environment was insufficient, then only partial loss of A_0 was observed, and it was quickly recovered when unmasked.
- Coordinated increase in fitness values of A_1-A_3, showing that that unmasking can bind together multiple abilities. This demonstration requires that the additional abilities can be optimised for acquiring vitamin C beyond the fitness required in their original function.

These two components can be predicted reliably from the amount of selection pressure in each stage of the simulation. Masking is due to a loss of selection pressure when the environmental source of vitamin C enables other abilities to acquire the daily limit of vitamin C. Unmasking by lower levels of vitamin C in the environment increases the selection pressure to acquire as much of the environmental source as possible, and hence increase the fitnesses of A_1-A_3.

What can simulations such as Hinton and Nowlan's and our simulations show? Like Hinton and Nowlan's simulation, the evidence that we present here is the internal computational consistency of the arguments. We can also compare the simulations, and study under what condition evolution's arrow causes genetic

assimilation, and under what conditions specificity is lost and genetic redistribution results.

The Baldwin effect shows how learning can guide evolution. The genetic redistribution effect shows how masking and unmasking by environmental changes can send the arrow of evolutionary change from specific genetic abilities to distributed suites of abilities, and binds those abilities into evolutionarily-synchronised cognitive components. Both can be viewed as transient dynamics in the strange loops from phenotype to genotype, stimulated by different environmental conditions.

1.5. Implications for cognitive science

It has been hypothesized that the Baldwin effect is a potential mechanism for evolving specific language adaptations for Universal Grammar and other modular capacities:

> *"The Baldwin effect probably played a large role in the evolution of brains. Contrary to the standard social science assumptions, learning is not some pinnacle of evolution attained only recently by humans. All but the simplest animals learn ... If the ability to learn was in place in an early ancestor of the multicellular animals, it could have guided the evolution of nervous systems toward their specialised circuits even when the circuits are so intricate that natural selection could not have found them on its own."*
>
> *Pinker (1997:179)*

From a cognitive modelling perspective, both the Baldwin effect and the conditions under which it occurs are important. However, the existence of learning alone does not determine whether a Baldwin effect will occur. Other conditions must also be satisfied. The evolution of specialised circuits will only benefit from the transient dynamic when the population is poised on the brink of discovery and the benefits and costs of learning are appropriately balanced.

Few people would argue whether the ability to learn language is in the genes. Clearly genes play a part. The argument in psycholinguistics is whether there are *specific* grammar genes, or whether there is a distributed learning system that is used for many cognitive abilities, not just learning languages. The psycholinguistic evidence is beyond this paper but the issue illustrates why the interactions between evolution and learning play a role in cognitive science.

Interestingly, Hinton and Nowlan's simulations of the Baldwin effect do not preclude the evolution of distributed abilities, as shown in the simulations in this study. Applied to the language argument, our simulations of the genetic redistribution effect support such a distributed evolutionary scenario. Deacon has argued that symbolic communication is a powerful masking agent that should transfer genetically specified abilities to distributed suites of abilities, and bind those abilities into

evolutionarily-synchronised cognitive components. The simulations in this paper demonstrate how such an effect could occur.

In conclusion, evolutionary computation is not a methodology that can resolve the issue of whether language is based on distributed or specific abilities. What it can show is that the underlying phenomena of "genetic assimilation" and "genetic redistribution" are both computationally coherent, and simulations can be used to investigate the conditions under which - *in silico* – the transient dynamics occur.

References

Baldwin, J. M. (1896). A new factor in evolution. *American Naturalist* 30: 441 – 451. Reproduced in (eds.) Belew, R.K. & Mitchell, M.,. *Adaptive Individuals in Evolving Populations*, Proceedings Volume XXVI, Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, Reading, MA.

Cohen, J. and Stewart, I. (1994). *The Collapse of Chaos*. London: Penguin.

Darwin, C. (1859). *The Origin of Species by Means of Natural Selection*. London: Murray.

Deacon, T. (2001). Presentation at the Conference on Language, Brain and Culture, Sydney, December.

French, R.M. and Messinger, A. (1994). Genes, *Phenes and the Baldwin Effect: Learning and Evolution in a Simulated Population*. In R. Brooks & P. Maes (Eds.), Artificial Life IV. Cambridge, MA: MIT Press.

Harvey, I. (1993). The puzzle of the persistent question marks: A case study of genetic drift. In S. Forrest, (Ed.) *Proceedings of the Fifth Int. Conf. on Genetic Algorithms*, San Mateo: Morgan Kaufmann.

Hinton, G. E. and Nowlan, S. J. (1987). How learning can guide evolution. *Complex Systems* 1: 495 –502.

Mayley, G. (1996). The evolutionary cost of learning. In (Eds) P. Maes, M. J. Mataric, Jean-Arcady Meyer, J. Pollack and S. W. Wilson, *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*. Cambridge, Mass: MIT Press, 458-467.

Morgan, C. L. (1896). On modification and variation. *Science*, 99:733-740.

Pinker, S. (1997). *How the Mind Works*. London: Penguin.

Turney, P., Whitley, D., and Anderson, R. (1996). *Evolution, Learning and Instinct: 100 Years of the Baldwin Effect*, (Eds). Cambridge, MA: MIT Press.

Watson, J., Geard, N. and Wiles, J. Genetic assimilation on an NK landscape. Submitted.

Watson, J. and Wiles, J. (2002). The rise and fall of learning: A neural network model of the genetic assimilation of acquired traits. To appear in the *Proceedings of the World Congress on Computational Intelligence*, May, 2002.

Wiles, J.; Schulz, R.; Bolland, S.; Tonkes, B.; and Hallinan,J. 2001. Selection procedures for module discovery: Exploring evolutionary algorithms for cognitive science. In Moore, J. D., and Stenning, K., eds., *Proceedings of the 23rd Annual Conference of the Cognitive Science Society (CogSci 2001),* 1124–1129. Lawrence Erlbaum Associates.