How Does Niche Construction Reverse the Baldwin Effect?

Hajime Yamauchi

Japan Advanced Institute of Science and Technology
`hoplite@jaist.ac.jp`

Abstract. Deacon [1] considers that the reverse Baldwin effect can be one of the major forces in language evolution. The reverse Baldwin effect is essentially a redistributional process of genes as a result of environmental changes which mask and unmask selection pressures. Although Deacon indicates that in the case of language evolution, niche construction is deeply involved in masking and unmasking processes, neither specific explanations for the mechanism nor examples have been given. In this study we use evolutionary computation simulations to explore how niche constructing properties of language evolution can induce at least the masking effect, and hence lead to genetic degradation. The simulation demonstrates that the masking effect is indeed a part of the evolutionary process found in the normal Baldwin effect.

1 Introduction

As a causal theory of learning and evolution, the Baldwin effect has gathered wide attention in evolutionary linguistics where first language acquisition is recognized as one of the key issues to understand this uniquely human capacity. However, for those accustomed to the Baldwinian view of the relationship between learning and evolution, what Deacon [1] describes may sound somewhat counterintuitive. While the Baldwin effect describes how previously learnt knowledge becomes a part of innate knowledge, according to Deacon, under some circumstances, innate knowledge would be replaced by more plastic, learnt knowledge. As the process seemingly follows the opposite flow of what the Baldwin effect describes, he called this process the "reverse Baldwin effect". This paper will present how the niche constructing aspect of language evolution serves as one of the key mechanisms necessary for the purported effect without assuming, as Deacon has, that externally motivated changes (like climate changes) in environmental conditions would take place.

2 Genetic Redistribution

While Waddington's [2] genetic assimilation is often conceived of as the mechanism of the Baldwin effect, Deacon believes that the reverse Baldwin effect is essentially a process of "genetic redistribution" where initially high innate

knowledge is functionally replaced by a suite of traits each of which has its own original function.

It is thought that genetic redistribution is triggered when the current selective pressure is somehow shielded. This is called the "masking effect" and it effectively works to reduce the adaptive importance of the trait carried. Consequently, the masking effect causes the gene expressing the trait to randomly drift. Eventually the gene loses its functional significance and becomes a "pseudogene"; incapable of expressing the given trait. When the mask is lifted (i.e. the original selective pressure resurfaces), the organism has to take an alternative option to compensate for the lack of the trait. Deacon posits that this "unmasking effect" causes originally irrelevant traits that evolved for different functions to form a suite in order to replicate the function of the original trait (and the gene for it).

Because the focal function is now subserved by the complexes of different traits attributed to a group of genes, the original selective pressure is now *redistributed* onto those genes. As the function is no longer controlled by a single gene, but a suite of different genes, Deacon considers it effectively increases the freedom of ontogenetic accommodation; the reverse Baldwin effect.

2.1 Case Studies

One of the empirical examples Deacon and others [1,3] often draw on is the case of vitamin C synthesis. Somehow, the ability of synthesizing ascorbic acid (vitamin C) was lost in the primate lineage including *Homo*. It is known that the gene responsible for the enzyme working on the last stage of the synthesis was replaced by its degraded pseudogene. Around the period when the gene irreparably deteriorated, it is assumed that the climate was warm allowing primates to obtain vitamin C rich fruits easily. This masked the adaptive importance of synthesizing the vitamin endogenously, and consequently the genes attributed to the function started to randomly drift. This effectively made the primates "dietarily entrenched" in a frugivorous life style. A change in climate, however, made the fruits scarcer, and unmasked the selective pressure once again. Because the gene responsible for the synthesis had become a pseudoegene because of the random drift, the primates had to rely on other traits to compensate. The alternative was to increase foraging ability to obtain fruits effectively. For that purpose, various primarily independent traits such as tooth structure, color vision, and taste preference were employed and increased their evolutionary importance: redistribution of selective pressure of the function originally served by the innate trait onto the suite of more genetically indirect traits. Wiles et al. [3] have conducted computer simulations based on the above argument, and confirmed that interactions between learning and evolutionary search could induce not only genetic assimilation, but also genetic redistribution if masking and unmasking effects take place.

Deacon (p.c.) has similarly attempted to explain the striking complexification of bird song during the domestication process of the whitebacked munia (*Lonchura striata* var. *domestica*) from its feral ancestor, the Bengalese finch

(*Lonchura striata*) [4]. Ritchie and Kirby [5] have demonstrated that this description of the song's complexification through the reverse Baldwin effect is computationally attestable.

2.2   Masking Effect in Language Evolution

Given the potential explanatory power of the reverse Baldwin effect, both Deacon [1] and Wiles et al. [3] envisage that it could also play a significant role in language evolution. However, it is apparent from the above arguments that, for the reverse Baldwin effect to take place, there needs to be some causal agent to induce the masking effect. In the case of vitamin C, it was the warm climate (and abundant fruits), and in the case of the munia, it was domestication. Deacon considers the potential masking agent in language evolution is its niche constructing process [6,7]. However, it is unclear quite how the process comes into play as regards the masking effect. In the next section, we will examine how complex interactions between language learning and evolution create niche constructing processes, and how they transfer the selective pressure on innate linguistic knowledge to learning, and hence lead to genetic drift.

3   A Computational Model

In order to establish if the niche constructing aspect of language could mask selective pressure, a computational simulation of language evolution has been developed. The simulation, which plausibly models both biological and cultural evolution, and the developmental aspect of language at a reasonable level, is based on my research [7] The model works with an evolving population of agents.

3.1   Stages

The main stages in the simulation are listed here, and details of each stage are given in following sections:

Birth The agent's grammar is empty regardless of innate linguistic knowledge coded in a chromosome *G-Chrom*, described in the next section. The agent's cognitive ability is determined based on her designated chromosome *L-Chrom*, also described in the next section.

Learning The agent is exposed to *nInput* linguistic inputs from her linguistic environment (i.e. E-language). *nInput* is considered to set the *critical period*. With the learning algorithm described in the next section, the agent builds her own grammar by using a particular cognitive ability *Cog*.

Invention The agent invents some parts of her grammar by using *Cog*.

Communication The agent communicates with her two immediate neighbor peers for *nCom* times. Successful communication will increase the agent's fitness score.

Reproduction Parents are selected probabilistically according to their fitness score and their chromosomes are crossed over using one-point cross-over to give two children. The two chromosomes (i.e. *G-Chrom* and *L-Chrom*) are independent, and cross-over takes place within each of the chromosomes. Individual genes are mutated with probability *pM*.

Death The entire grammar of each agent in the population is sampled and the resulting linguistic information creates an E-language from which the next generation is to learn.

3.2   Model Structure

1. **The Agent** An agent has two types of chromosomes, namely *G-Chrom* and *L-Chrom*. Genetic information of innate linguistic knowledge is coded in *G-Chrom*. The innate linguistic knowledge is encoded as a string of 0*s*, 1*s*, and ***s in *G-Chrom*. ***** is thought of a junk allele, and no use for anything. In the results reported here, the entire knowledge is represented in a 24 ternary string (i.e. the number of genes in *G-Chrom* is 24). The initial gene pool has random numbers of 0*s*, 1*s*, and *s (thus, on average, one agent has 8 of each allele).

    The agent also has a different type of genetic information on *L-Chrom* coding the size of the cognitive capacity *Cog*. *Cog* is a cognitive capacity which enables the agent to update her grammar when discrepancies exist between her innate linguistics knowledge and linguistic inputs, and invent her own grammatical information. The size of *Cog* is given by a quantifiable scalar value *nCog*. The value is determined based on the number of 1*s* in *L-Chrom*. *L-Chrom* consists of a 48 binary string. The incremental value *iV* is set to 1 in the reported simulation. Therefore, the maximum value of *nCog* is 48.

    A grammar is coded as a ternary string, and the length of the string is 24 –equal to the size of *G-Chrom*. Three possible allelic values are 0, 1 and *NULL*. Wherever there is a *NULL* allele on the grammar, this part of the grammar is considered NOT to code any linguistic knowledge. Therefore, the more the *NULL* alleles there are in a grammar, the less the grammar codes linguistic knowledge. As described in the next section, although the agent learns a particular grammar by using her innate linguistic knowledge and input data, the grammar does not necessarily faithfully reflect the innate linguistic knowledge coded in *G-Chrom*; some parts of the grammar may converge to different values from the corresponding parts of the innate knowledge.

2. **Learning** Every agent in every generation is born with a completely empty grammar; all 24 alleles are *NULL*. Learning is the process to update such *NULL* alleles to substantial alleles (i.e. 0*s* and 1*s*). A learning agent sequentially receives linguistic inputs from *nAdult* adult neighbors (this is set to 5 in the simulation). Adults are the agents from the previous generation. A linguistic input is thought of as the *utterance* of an adult, which is represented by one allele of her mature grammar. Utterances derived from *NULL* alle-

les are considered as *NULL* utterances, and no learning (thus no grammar update) takes place. Following is the algorithm to develop the grammar:

Learning Algorithm Whenever the learner receives a linguistic input:

1. If the input value and the allelic value of the corresponding locus of the learner's grammar are different(i.e. not "parsable"), carry out the following the procedures:

   (a) If the corresponding allele of *G-Chrom* (her innate linguistic knowledge) "matches" (i.e. the two values are the same) with the input, update the given allele of the current grammar, and subtract 1 point from *nCog*.

   (b) If the corresponding allele of the innate linguistic knowledge is different from the input, update the given allele of the current grammar, and subtract 2 points from *nCog*.

2. Otherwise keep the current grammar.

The learning procedure is ceased when either *nCog* reaches 0 or the number of inputs meets *nInput*. *nInput* is set to 120. *NULL* utterances are counted for this process. Any locus of the grammar not receiving any input (or receiving only *NULL* utterances) remains *NULL*. Who utters and which part of her grammar is provided an input datum is totally random. This means that if the adults have totally different grammars, the learner may update a given allele of her grammar frequently.

3. **Invention** Agents are capable of inventing their own linguistic knowledge, and of adding it onto their grammars. If an agent still holds *NULL* alleles in her grammar after the learning has taken place, and if her *nCog* has not yet become 0, with a probability of *pI*, pick one *NULL* allele randomly, and flip it to either 0 or 1 randomly, and subtract 1 point from *nCog*. This process is carried on until either no more *NULL* allele is in the grammar, or *nCog* reaches 0. In reported the simulation, *pI* is set to 0.02. Once the invention process is over, her grammar is considered to have reached a mature state, and no more grammar update takes place.

4. **Communication** Each agent is involved in *nCom* communicative acts with her immediate peer neighbors. The fitness of an agent is scored on the basis of how many utterances spoken with her mature grammar were parsable to the hearer[1]. The same as for learning input, an utterance is represented by one allele of the speaker's mature grammar. *nCom* is set to 15 for utterances. As each neighbor also speaks to each agent the same number of times, a total of 30 communicative acts is involved to gauge her fitness. Each successful communication increases her fitness by 1. Those who cannot establish any communication still receive 1 fitness score to keep the possibility of being a parent in Reproduction. Therefore, the maximum fitness value is 31.

5. **Reproduction** Roulette wheel selection is used for selecting parents according to their fitness, and their chromosomes *G-Chrom* and *L-Chrom* are individually crossed over using single-point crossover to create two offspring. *pM* is set to 0.001 per allele.

---

[1] By the configuration of the model, this also means that how many utterances of the other can be parsed with her grammar

3.3   Layout of the Model

In this section, the overall layout of the model is described. The layout basically follows Kirby and Hurford [8]. Fig. 1 shows how the elements of the simulation are organized.
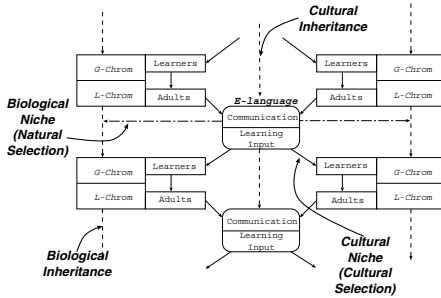

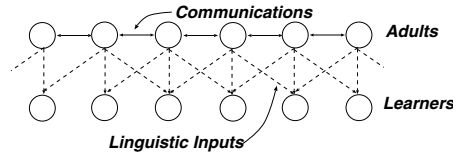**Fig. 1.** An overview of the simulation model.

Fig. 2. The spatial organization of the population.

The figure shows how the niche constructing properties of the language are modeled in the simulation. Notice that there are two different channels of inheritance that interact with each other; namely cultural and biological inheritances. Linguistic knowledge is inherited in the biological channel via genes (i.e. innate linguistic knowledge) and is also inherited in the cultural channel via learning. What is inherited through the cultural channel is *niche constructed* grammars; accumulations of previous generations' linguistic inventions become the learning environment from which learners *reconstruct* their ancestors' grammars. The other side of the coin is biological niche construction; grammars acquired with that manner also affect the natural selection of individuals. The adaptive utility of a given grammar is only determined when a specific linguistic demography of the population is referred[2]. As such a demography is created by previous generations' linguistic activities (including both learning and communications), it is also a product of niche construction; in this case, what is constructed is a biological niche which affects individuals fitness.

Fig. 2 shows the spatial organization of the populations. Individuals are organized in a one-dimensional loop. Circles on the two different tiers represent adults and learners (i.e. two different generations). Incidents of communications (represented by the horizontal double-arrows in the adult tier) only take place within a single tier and is local since an individual attempts to communicate with her two immediate neighbor peers (left and right, in this figure). While this is an adult-to-adult process that results in natural selection, learning is thought of a vertical, adult-to-child transmission which results in cultural inheritance. One

---

[2] A particular grammar is only useful when others communicate with the same or very similar grammars.

adult provides linguistic inputs for *nAdult* neighbor learners (from the learner's point of view, she receives the inputs from *nAdult* immediate neighbor adults).

Note that reproduction is not affected by this spatial organization. Parents are probabilistically selected in proportion to their fitness and their offspring is randomly distributed in this space.

4   Results

All figures shown here are taken from one typical run of the simulation under the conditions described, and as such they well characterize the general tendency of the model. Fig.3 shows the average fitness of the population over time with a red line, and the average number of *NULL* alleles in matured grammars with a blue line. Rapid increase of the fitness (reaching 30 around the 150th generation) shows the population quickly evolves to almost the optimal state (the highest possible score is 31) as they develop their linguistic knowledge (i.e. reduction of *NULL* alleles). They do not do so randomly, rather neighbor individuals develop similar grammars in order to establish communications with peers.

Fig.4 shows the average number of * ("Junk") alleles in *G-Chrom* in a red line, the average matching number *nMatch* of matured grammars and *G-Chrom*, and evolution *Cog* by presenting *nCog* in a green line. We compare learners' stable grammar (i.e. after the learning, but before the invention stages) with their *G-Chrom* (i.e. innate linguistic knowledge. *nMatch* indicates a degree of contribution of innate linguistic knowledge to a given mature grammar. The higher the matching number is, the more of one's innate knowledge is reflected on the matured grammar. From the graph, it is apparent that soon after *Cog* has evolved, genes in *G-Chrom* start to match with grammars, and rapidly hit the
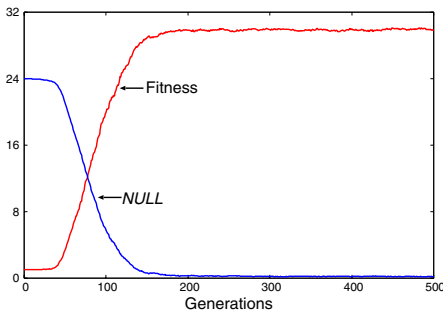

Fig. 3. Average fitness and numbers of *NULL* alleles in grammars of the population against generations. Because of these factors evolved rapidly, the graph shows only the first 500 generations.
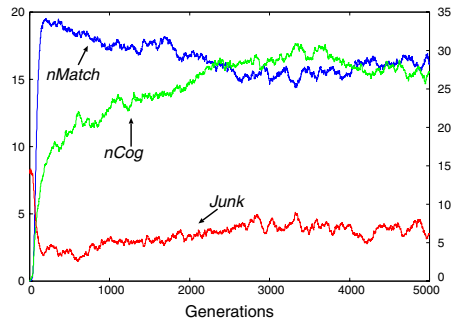
Fig. 4. In this graph, average numbers of "Junk" alleles ("*" alleles), *nMatch*, and *nCog* are shown. Values of *nCog* are plotted against the y2 axis (right).

highest point (on average 20 genes match with alleles in grammars at around the 200th generation) while $nCog$ is still able to override only 14 alleles. The average number of the junk allele also supports this view. As $nMatch$ grows, the junk allele rapidly expelled from the population. However, as Cog gradually evolves, $nMatch$ decreases and becomes stable around 15. Along the same line, the junk allele also kicks in the population once again.

5   Analysis

The time course of the simulation demonstrates a complex interactions of learning, its evolution, and evolution of innate linguistic knowledge. Overall, we are able to summarize the course in three main stages: Stage 1: *Baldwinian Niche Construction* 0-200 (generations); Stage 2: *Masking* 200-4000; and Stage 3: *Stable* 4000-.

5.1   Baldwinian Niche Construction

The initial rapid increase of fitness in Stage 1 goes hand in hand with the increase of genetic contribution in language acquisition. What is striking is that this trend starts immediately after the average $nCog$ reaches 1 (around the 30th generation) and it reaches almost the highest possible value, while $nCog$ still remains around 12 which barely covers a half of the whole grammar to be updated. This strongly implies that innate linguistic knowledge of learners highly faithfully represents invented grammars of previous generations. The population at the first few generations would not receive any input from the previous generations as with 0 Cog, neither learning nor invention takes place to leave input for next generation. However, soon after Cog evolves in some individuals, they can invent their own grammatical information randomly, they do not use Cog for learning (because previous generations neither learn nor invent their grammars) and can utilize it to invent their own grammar. Their grammar would be just a fraction of the full-size grammars, but those agents can leave the grammars for the next generation's linguistic input. For the next generation, the potentially most adaptive individuals would be those who have genetic alleles matching with linguistic inputs, and spare Cog for inventing further linguistic knowledge: with extremely limited Cog, individuals who have to learn (update) their grammar would completely lose the chance to push the envelop of their grammar, and fail to establish communication. Then some of the individuals in the next generation have genes matching with further input utterances, and repeat the same process. This is a cyclic process which Avital and Jablonka [9] have called "assimilate stretch": learning pushes the envelop ⇒ genes assimilate it ⇒ leave the room for the learning to push the envelop further. Consequently, * alleles in *G-Chrom* are quickly winnowed out, and 0s and 1s are reshuffled to match the given linguistic environment. Importantly, this cycle of adaptive assimilations is the consequence, and the cause of the niche construction process in the linguistic environment: what is created here is linguistic knowledge which works as a

niche for both cultural inheritance and biological inheritance: individuals learn from linguistic knowledge created (invented) by the previous generations (niche construction in the cultural channel), and adaptability of their learnt grammars is evaluated only within local linguistic communities (through communications with neighbor peers). The assimilated innate linguistic knowledge also canalizes what type of grammars are learnable, and it reflects on agents' learnt grammars. Those grammars determines the shape of the selective environment (the selection is frequency dependent as the utility of a given grammar is determined by the frequency of other grammars in neighbors). This is what I [7] report as Baldwinian Niche Construction.

5.2   The Masking Effect

Once niche constructed linguistic knowledge, which is a cooperative product of highly limited *Cog* and *G-Chrom*, has been transferred into innate linguistic knowledge, the masking effect is able to take place. As *Cog* evolves in Stage 2, the importance of the contribution of innate linguistic knowledge begins to be masked. This permits random drift to affect genes in *G-Chrom*: as learning can cover the discrepancy between inputs and innate knowledge, the number of the junk allele gradually increases. The result is a coordinated decrease of *nMatch*. This trend continues until around the 3000th generation, where *G-Chrom* and *nMatch*, and the number of the junk allele reach their (somewhat rough) plateaus (Stage 3). During Stage 2, *nMatch* decreases by nearly 5, and remains around 15 in Stage 3. Therefore, once again *Cog* shoulders the important role of language acquisition.

It is highly significant that, given the early increase of fitness, the selective pressure for the increase of *Cog* seems to be well diminished at the end of Stage 1, while in fact it keeps evolving. This is potentially because a different type of random drift takes place in *L-Chrom*: initially the chromosome does not include any 1*s*. Because of mutations in the genes, however, gradually the numbers of the alleles increase, and spreads by crossovers. In other words, the evolution of *Cog* is a largely neutral one. This indicates that the masking effect is an epiphenomic consequence of the neutral evolution.

6   Discussion and Conclusion

The simulation clearly demonstrates the self-induced model of the masking effect by the niche constructing property of language. The result is comparable to those provided in Wiles et al. [3], and Ritchie and Kirby [5]. However, both models employ arbitrary changes in environmental conditions to induce the masking effect. In this regard, our model successfully presents one possible mechanism of the niche construction based masking effect as Deacon envisages it.

As it stands though, the model does not explain how and why the unmasking effect would occur. The model only shows that a certain degree of the masking effect takes place; this does not cause a substantial degradation of genetic

information which is necessary for the unmasking effect (although our model fundamentally lacks the mechanism to induce the unmasking effect due to the way that *G-Chrom* and *L-Chrom* interact). We observe that if *Cog* is arbitrarily knocked down (i.e. reduce *nCog*) after it reaches the plateau, the genetic contribution could manage to recover its strong effect on learning again. However, as Wiles et al. [3] state, unmasking may not require explicit environmental changes to necessary conditions for genetic redistribution to take place.

Also, the model does not demonstrate genetic redistribution. This is because our model design lacks distributional properties in *L-Chrom*. Rather it encodes a type of cognitive ability in a highly simplified manner. However, we think that genetic redistribution is one of the *consequence*s of masking and unmasking effects, but not a *mechanism* of it. Indeed, by modifying the model design, we consider that we would be able to demonstrate a type of redistributional process. Despite that, in this simulation we have intentionally ignored that aspect since it would potentially blur the causal mechanism working between niche constructions and the masking effect. It remains, however, one of our next research targets in this avenue.

References

1. Deacon, T.W.: Multilevel selection in a complex adaptive system: The problem of language origins. In: Weber, B.H., Depew, D.J. (eds.) Evolution and Learning, pp. 81–106. MIT Press, Cambridge (2003)
2. Waddington, C.H.: Genetic assimilation of an acquired character. Evolution 7, 118–126 (1953)
3. Wiles, J., Watson, J., Tonkes, B., Deacon, T.: Transient phenomena in learning and evolution: Genetic assimilation and genetic redistribution. Artificial Life 11(1-2), 177–188 (2005)
4. Okanoya, K.: Sexual display as a syntactic vehicle: The evolution of syntax in birdsong and human language through sexual selection. In: Wray, A. (ed.) The Transition to Language, Oxford University Press, Oxford (2002)
5. Ritchie, G., Kirby, S.: Selection, domestication, and the emergence of learned communication systems. In: Second International Symposium on the Emergence and Evolution of Linguistic Communication (2005)
6. Odling-Smee, F.J., Laland, K.N., Feldman, M.W.: Niche Construction: The Neglected Process in Evolution. Monographs in Population Biology, vol. 37. Princeton University Press, Princeton (2003)
7. Yamauchi, H.: Baldwinian Accounts of Language Evolution. PhD thesis, The University of Edinburgh, Edinburgh, Scotland (2004)
8. Kirby, S., Hurford, J.: Learning, culture and evolution in the origin of linguistic constraints. In: Husbands, P., Harvey, I. (eds.) Proceedings of the Fourth European Conference on Artificial Life (ECAL-97), pp. 493–502. The MIT Press, Cambridge (1997)
9. Avital, E., Jablonka, E.: Animal Traditions: Behavioural Inheritance in Evolution. Cambridge University Press, Cambridge (2000)