# John Benjamins Publishing Company

# Sequential learning and the interaction between biological and linguistic adaptation in language evolution

Florencia Reali and Morten H. Christiansen
Department of Psychology, Cornell University, Ithaca, NY 14853

It is widely assumed that language in some form or other originated by piggy-backing on pre-existing learning mechanism not dedicated to language. Using evolutionary connectionist simulations, we explore the implications of such assumptions by determining the effect of constraints derived from an earlier evolved mechanism for sequential learning on the interaction between biological and linguistic adaptation across generations of language learners. Artificial neural networks were initially allowed to evolve "biologically" to improve their sequential learning abilities, after which language was introduced into the population. We compared the relative contribution of biological and linguistic adaptation by allowing both networks and language to change over time. The simulation results support two main conclusions: First, over generations, a consistent head-ordering emerged due to linguistic adaptation. This is consistent with previous studies suggesting that some apparently arbitrary aspects of linguistic structure may arise from cognitive constraints on sequential learning. Second, when networks were selected to maintain a good level of performance on the sequential learning task, language learnability is significantly improved by linguistic adaptation but not by biological adaptation. Indeed, the pressure toward maintaining a high level of sequential learning performance prevented biological assimilation of linguistic-specific knowledge from occurring.

## 1. Introduction

Although the space of logically possible languages is vast, the world's languages only take up a small fraction of it. As a result, human languages are characterized by a number of universal constraints on how they are structured and used. Many of these constraints undoubtedly derive from innate properties of the learning and processing mechanisms brought to bear on language acquisition and processing. But what is the origin of these constraints in our species?

One approach suggests that language evolved through a gradual process of natural selection of more and more complex linguistic abilities (e.g., Briscoe, 2003; Dunbar, 2003; Jackendoff, 2002; Nowak, Komarova & Nyogi, 2002; Pinker, 1994, 2003; Pinker & Bloom, 1990). From this perspective, biological adaptation has endowed humans with a large body of innate knowledge specific to language: A Universal Grammar. Supported by a rapidly growing bulk of research from linguistics (grammaticalization: Givón, 1998; Heine & Kuteva, 2002), archeology (Davidson, 2003), the development of indigenous sign-languages (Ragir, 2002), and computational modeling (e.g., Batali, 1998; Kirby, 2001 — see Kirby, 2002, for a review), an alternative perspective has emerged, focusing on the adaptation of language itself — *linguistic adaptation* — rather than on the adaptation of biological structures such as the brain. On this account, linguistic adaptation resulting from cultural transmission of language across many generations of language learners has resulted in the emergence of complex linguistic structure (e.g., Christiansen, 1994; Christiansen & Chater, 2008; Deacon, 1997; Kirby & Hurford, 2002; Tomasello, 2003). The universal constraints we observe across the world's languages are proposed to be a consequence of the process of cultural transmission combined with cognitive limitations on learning and processing (Kirby & Christiansen, 2003; see Christiansen & Chater, 2008, for a review).

Cultural transmission, however, does not take place in a vacuum but within the broader context of the biological evolution of the hominid species. A complete picture of the role of cultural transmission in language evolution must therefore take into account the complex interplay between general biological adaptation and linguistic adaptation. Recent computational studies have explored the role of biological adaptation for language (e.g., Batali, 1994; Cangelosi, 1999; Nowak et al., 2002) and linguistic adaptation (e.g., Batali, 1998; Kirby, 2001). Moreover, a growing number of studies have started to investigate the potentially important interactions between biological and linguistic adaptation in language evolution (Christiansen, Reali & Chater, 2006; Hurford, 1989; Hurford & Kirby, 1999; Kvasnicka & Pospichal, 1999; Livingstone & Fyfe, 2000; Munroe & Cangelosi, 2002; Smith 2002; 2004; Yamauchi, 2001).

However, the complex interactions between biological and linguistic adaptation are also subject to further limiting factors, deriving from the constraints on the neural mechanisms that are used to learn and process language (Christiansen & Chater, 2008) as well as the social context within which language is acquired and used (Levinson, 2000). In this paper, we conduct evolutionary simulations to further explore how these interactions may be affected by the first type of constraints arising from the brains of the language learners, focusing on how the important cognitive ability of sequential learning may influence the evolution of language structure. Two main results are reported. First, we provide evidence suggesting

that apparent 'arbitrary' aspects of linguistic structure – such as word order universals – may arise as a result of sequential learning and processing constraints. Consistent with previous studies (e.g., Christiansen & Devlin, 1997; Kirby, 1998), our simulations revealed that consistent head-ordering emerged over generations of evolving learners and languages as a result of linguistic adaptation. Second, we explore the interaction between sequential learning constraints and biological adaptation. We assume that after the emergence of language, sequential learning skills would still have been crucial for hominid survival. Thus, the simulations were designed to explore the relative contribution of linguistic and biological adaptation while simulating a selective pressure toward maintaining non-linguistic sequential learning abilities. The simulations revealed that, under such conditions, language learnability is significantly improved by linguistic adaptation but not by biological adaptation. Indeed, the pressure toward maintaining a high level of sequential learning performance prevented biological adaptation from occurring.

## 2.   Sequential learning and language evolution

There is an obvious connection between sequential learning and language: Both involve the extraction and further processing of elements occurring in temporal sequences. Indeed, recent neuroimaging and neuropsychological studies point to an overlap in neural mechanisms for processing language and complex sequential structure. A growing bulk of work indicates that language acquisition and processing shares mechanisms with sequential learning in other cognitive domains (e.g., language and musical sequences: Koelsch et al., 2002; Maess, Koelsch, Gunter & Friederici, 2001; Patel, 2003, Patel, Gibson, Ratner, Besson & Holcomb, 1998; sequential learning in the form of artificial language learning: Christiansen, Conway & Onnis, 2007; Friederici, Steinhauer & Pfeifer, 2002; Petersson, Forkstam & Ingvar, 2004; break-down of sequential learning in aphasia: Christiansen, Kelly, Shillcock & Greenfield, 2007; Hoen et al., 2003). For example, using event-related potential (ERP) techniques, Friederici et al. (2002) showed that subjects trained on an artificial language have the same brainwave patterns to ungrammatical sentences from this language as to ungrammatical natural language sentences (see also Christiansen et al., 2007). In a different series of studies, Patel et al. (1998), showed that novel incongruent musical sequences elicit ERP patterns that are statistically indistinguishable from syntactic incongruities in language. Using event-related functional magnetic resonance imaging (fMRI) methods Petersson et al. (2004) have shown that Broca's area, which is well-known for its involvement in language, is also active in an artificial grammar learning tasks. Moreover, results from a magnetoencephalography (MEG) experiment further suggest that Broca's

area is involved in the processing of music sequences (Maess et al., 2001). Together, these studies suggest that the same neural mechanisms that underlie processing of linguistic structure are involved in non-linguistic sequential learning.

Here we argue that this close connection is not coincidental but came about because the evolution of our linguistic abilities to a large extent has "piggybacked" on sequential learning and processing mechanisms existing prior to the emergence of language. Human sequential learning appears to be more complex (e.g., involving hierarchical learning) than what has been observed in non-human primates (Conway & Christiansen, 2001). As such, sequential learning has evolved to form a crucial component of the cognitive abilities that allowed early humans to negotiate their physical and social world successfully. Constraints on sequential learning would then, over hundreds of generations, have shaped the structure of language through linguistic adaptation, thus giving rise to many linguistic universals (Bybee, 2002; Christiansen, Dale, Ellefson & Conway, 2002; Ellefson & Christiansen, 2000). On this account, language could not have "taken over" these learning mechanisms because the ability to deal with sequential information in the physical and social environment would still have been essential for survival (as it is today — see Botvinick & Plaut, 2004, for a review).

The approach favoring biological adaptation also relies on pre-existing learning mechanisms to explain the initial emergence of language. For example, Pinker and Bloom (1990) speculated that, "(…) the multiplicity of human languages is in part a consequence of learning mechanisms *existing prior* to (…) the mechanisms specifically dedicated to language" (p. 723; our emphasis). Through biological adaptation, these learning mechanisms would then gradually have become dedicated to language, incorporating innate linguistic knowledge. The evolutionary mechanism by which language principles are proposed to have become genetically encoded through gradual assimilation is known as the Baldwin effect (Baldwin, 1896; Waddington, 1940 — see also contributions in Weber & Depew, 2003). Although a Darwinian mechanism, the Baldwin effect resembles Lamarckian inheritance of acquired characteristics in that traits that are learned or developed over the life span of an individual become gradually encoded in the genome over many generations. Biological adaptation for language via the Baldwin effect (e.g., Briscoe, 2003; Pinker, 1994; Pinker & Bloom, 1990) can be summarized in the following steps:

1.  Initially language feature F is *learned* from exposure to a language in which F holds
2.  Genes that make learning F faster are selected
3.  Eventually, F may be known with no experience
4.  F is coded genetically

The Baldwin effect so construed may not only help explain how biological adaptations for language could gradually emerge, but it may also introduce a potential caveat for the cultural-transmission approach to language evolution. It is possible to grant that many aspects of language structure could emerge as a consequence of linguistic adaptation, but then still argue that the resulting linguistic features would then subsequently gradually become innate due to the Baldwin effect. However, on the sequential-learning account presented here, the Baldwin effect would not cause the original learning mechanisms to become dedicated to language because the ability to deal with sequential information in the physical and social environment would still have been essential for survival. Nonetheless, we consider this to be an empirical issue that can be addressed by computational means, and to which we turn next.

The first set of computational simulations explores the interactions between linguistic and biological adaptation under constraints derived from sequential learning. In the second set of simulations we further explore the impact of the sequential learning constraints on language evolution. Recent computational work suggests that biological assimilation via Baldwin effect may not be possible when the target – *language* – changes over time (Chater, Reali & Christiansen, 2009; Christiansen, Reali & Chater, 2006). Simulation 2 was designed to show yet another caveat for the adaptationist view: Gradual assimilation of linguistic knowledge may not be feasible when the underlying neural machinery does have to accommodate other non-linguistic tasks. To test this hypothesis, in Simulation 2 we manipulated the presence/absence of sequential learning constraints. To establish the individual effect of this factor, we controlled for linguistic adaptation by keeping the language constant throughout the simulations. The results suggest that biological adaptation is possible when removing the pressure to maintain the networks' ability for sequential learning. However, sequential-learning constraints on their own are sufficient to counter the effects of biological adaptation toward language-specific knowledge. We conclude by discussing the further implications of our simulations for research on language evolution.

## 3.   Simulation 1: Biological vs. linguistic adaptation

There have been several computational explorations of the Baldwin effect (e.g., Briscoe, 2002; Hinton & Nowlan, 1987; Munroe & Cangelosi, 2002). Of most relevance to our simulations presented below is a study by Batali (1994), showing that it is possible to obtain the Baldwin effect using simple recurrent networks (SRNs; Elman, 1990) trained on context-free grammars. Over generations, network performance improved significantly due to the selection and procreation of

the best learners. In the present study, we adopt a similar approach but introduce different assumptions concerning the nature of the task and considering the effect of pre-linguistic sequential learning constraints.

Our simulations involved generations of 9 differently initialized SRNs. An SRN is essentially a standard feed-forward neural network equipped with an extra layer of so-called context units. At a particular time step *t* an input pattern is propagated through the hidden unit layer to the output layer. At the next time step, *t+1*, the activation of the hidden unit layer at time *t* is copied back to the context layer and paired with the current input. This means that the current state of the hidden units can influence the processing of subsequent inputs, providing a limited ability to deal with integrated sequences of input presented successively. This type of network is well suited for our simulations because they have previously been successfully applied both to the modeling of non-linguistic sequential learning (e.g., Botvinick & Plaut, 2004; Servan-Schreiber, Cleeremans & McClelland, 1991) and language processing (e.g., Christiansen, 1994; Christiansen & Chater, 1999; Elman, 1990, 1991).

In order to simulate the emergence of pre-linguistic sequential learning abilities, we first trained the networks on a learning task involving the prediction of the next element in random five number-digit sequences. We allowed the networks to evolve "biologically" by choosing the best network in each generation, permuting its initial weights slightly to create 8 offspring, and then training this new generation on the sequential learning task. After 500 generations the error on sequential learning was reduced considerably, and we introduced language into the population. Thus, the networks were now trained on both sequential learning and language. Crucially, both networks and language were allowed to evolve, so
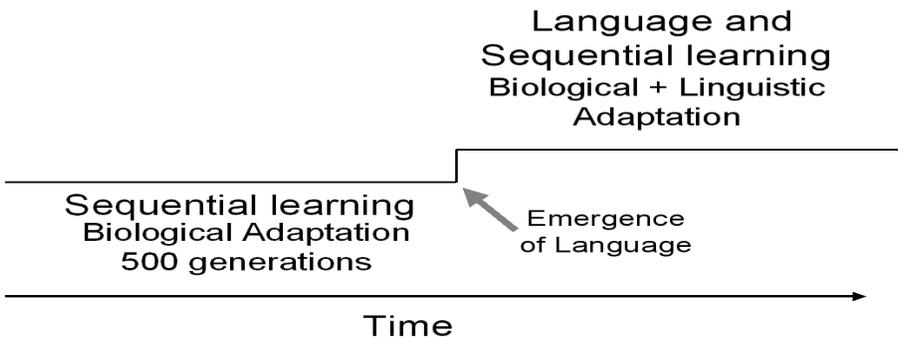


**Figure 1.**  A schematic outline of the simulation timeline. During the first 500 generations, the networks improve their sequential learning abilities through biological adaptation. Language is then introduced into the population. Both networks and languages are allowed to evolve to improve learning.

that we were able to compare the relative contribution of biological and linguistic adaptation. For each generation, we selected the networks that performed best at language learning with the additional constraint that they were also required to maintain their earlier evolved ability for sequential learning (on the assumption that this type of learning would still be as important for survival as it was prior to language). At the same time, linguistic adaptation was implemented by selecting the best-learnt language as the basis for the next generation of languages. Fig. 1 shows the basic timeline for the simulations.

### 3.1 Method

#### 3.1.1 *Networks*

Each generation in our simulations contained nine SRN learners. The networks consisted of 21 units in the input layer, 6 units in the output layer and 10 units in the hidden and context layer. The initial weights of the first generation of networks were randomly distributed uniformly between −1 and +1. Learning rate was set to 0.1 with no momentum.
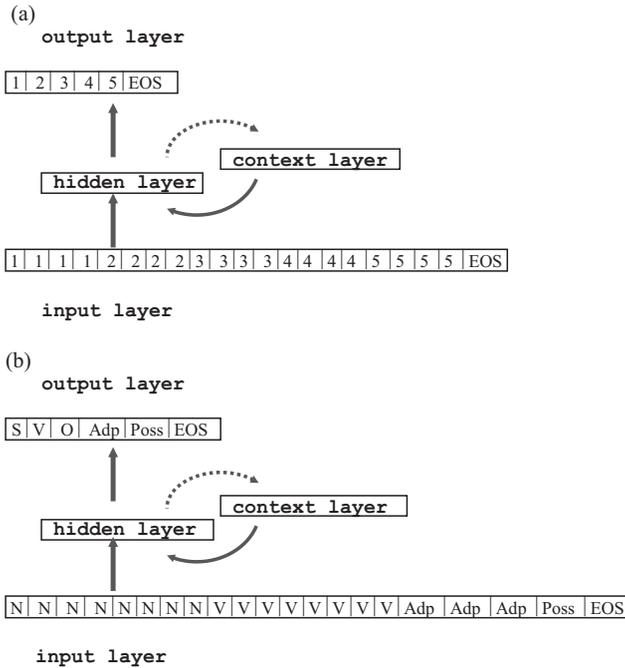


**Figure 2.** Network configuration for the sequential learning task (a) and linguistic task (b). The arrows indicate full connectivity between layers. Dashed lines indicate fixed connection weights (with a value of 1), and solid lines indicate learnable connection weights.

Networks trained on the sequential learning task had a localist representation of digits. In the input layer, four units represented each digit, however, each time a digit was presented to the network, only one unit was active at a time with equal probability.[1] Additionally, one input unit represented the end of the string (EOS). Each unit in the output layer represented a digit from 1 to 5 and one unit representing EOS. Fig. 2a provides an illustration of the sequential-learning configuration of the SRN.

When networks were trained on the linguistic task, each input to the network contained a localist representation of the incoming word: Each unit represented a different word in the vocabulary (20 total) and one unit represented the end of sentence (EOS). In the output layer each unit represented a grammatical category/thematic role — subject (S), verb (V), object (O), adposition (Adp), and possessive (Poss) — and one unit represented EOS. The SRN configuration for the language-learning task is shown in Fig. 2b. Networks were trained using the backpropagation algorithm.

### 3.1.2   *Materials*

*Sequential learning task.* For our sequential-learning simulations, we used a modified version of a serial reaction-time task, originally developed by Lee (1997) to study implicit learning in humans, and previously simulated using SRNs (Boyer, Destrebecqz & Cleeremans, 1998). The task requires predicting the next digit in a five-digit string. Digits went from 1 through 5 and were presented in a random order. However, the following simple rule constrained possible sequences of digits: Each of the five different digits can only appear once in the string. For instance, the sequence "34521" is legal, while the sequence "34214" is not. Therefore, the underlying rule is a gradient of probabilities across the five positions, where the first digit in the sequence is completely unpredictable and the last one is completely predictable. This task is particularly challenging because the information required to predict the last digit in the sequence goes beyond the information conveyed in transitional probabilities of co-occurrence of pairs or triples of digits. In order to predict the last digit, the network needs to keep track of the previous four positions.

*Language and linguistic task.* The languages were generated by phrase-structure grammars, defined by a system of rewrite rules determining how sentences are constructed. The phrase-structure grammar "skeleton" used in this simulation is presented in Fig. 3a, comprising six rewrite rules involving the following major constituents: sentence (S), verb phrase (VP), noun phrase (NP), adpositional phrase (AP), and possessive phrase (PossP).[2] Individual grammars contained variations in the head order of each rewrite rule, varying among three possible values: head first, head last, and flexible head order. In order to simulate language

variation, head order was modified by shifting the constituent order of a rewrite rule. For example, a grammar with the rule NP → N (AP), a head first rule, could be made head final by simply rewriting NP as (AP) N, with the head of the noun phrase in the final position. Alternatively, if the rewrite rule has flexible head order, the phrase is rewritten as head first or head final with equal probabilities in a sentence. Fig. 3b provides an example of an instantiated grammar defined by a particular head order arrangement. All possible combinations of head order in the six rewrite rules define the space of all possible grammars ($3^6 = 729$).

Networks were trained using a simple vocabulary consisting of 20 words: 8 nouns, 8 verbs, 3 adpositions and 1 possessive marker. Each word in the input was mapped on to one of the following five grammatical roles: Subject, Verb, Object, Adposition and Possessive. The networks' task was to predict the next grammatical role in the sentence. Successful network learning thus required sensitivity to grammatical role assignments, allowing us to compare the ease with which the SRN was able to learn the majority of the fixed orders of subject (S), verb (V) and object (O): SOV, SVO, VOS, and OVS (accounting for nearly 90% of language types, Van Everbroeck, 1999).

(a) **Grammar Skeleton**        (b) **Grammar Example**

| | | | | | | |
|---|---|---|---|---|---|---|
| S | → {NP VP} | (1) | S | → NP VP \| VP NP | (Flex) |
| NP | → {N (AP)} | (2) | NP | → (AP) N | (HFinal) |
| AP | → {Adp NP} | (3) | AP | → Adp NP | (HFirst) |
| VP | → {V(NP)(PP)} | (4) | VP | → V(NP)(PP) | (HFirst) |
| NP | → {N PossP} | (5) | NP | → N PossP\|PossP N | (Flex) |
| PossP | → {Poss NP} | (6) | PossP | → NP Poss | (HFinal) |

**Figure 3.** a) Grammar skeleton: Curly brackets represent changeable head order and round brackets represent optional phrases. Probability of recursion is 1/3. b) Example of one possible grammar constituted by a particular head order combination of the six rewrite rules (Flex=Flexible rewrite rule; HFirst = head first; HFinal = head final).

### 3.1.3    *Procedure*

As indicated in Fig.1, the networks were initially trained on the sequential learning task and allowed to evolve biologically. During every generation each network was trained on 500 random strings of digits and tested on 100 strings. After 500 generations, language was introduced into the population and the networks were trained on both sequential learning and language. The weights were reset to their biologically-evolved initial settings between the two tasks, so that the network had identical starting conditions when learning sequential structure and language. This stage involved biological competition between nine networks and linguistic competition between five grammars. For each grammar, the networks were trained on the linguistic task using 1,000 sentences and were tested on 100 sentences. The

"best learner" network and "the best learnt" grammar in each generation were selected as the basis for the next generation, thus allowing us to pitch biological and linguistic adaptation against each other.

We measured performance on the sequential learning task by comparing network predictions with the ideal output (had the network learned the task perfectly). For each position in a sequence, we calculated the cosine[3] of the angle between the output vector (network predictions) and the theoretically derived probability vector for the next digit given the previous digit(s). The overall score for the sequential-learning task was then computed as the mean cosine across all positions in all test strings.

Performance on the linguistic task was scored by comparing network predictions for each grammatical role to the probabilistically ideal output given the previous words in the utterance. For each word, we compared the full conditional probability vector for the possible next grammatical role to the output vector representing the network predictions (see Christiansen & Chater, 1999, for details), calculating the cosine to the angle between the two vectors. The overall score for the language-learning task was then computed as the mean cosine across all words in all test utterances.

*Biological Adaptation.* We allowed the networks to evolve "biologically" by choosing the best network in each generation, permuting its initial weights slightly to create 8 offspring. In every generation, the networks were trained and the fitness assessed in terms of their performance on the linguistic and sequential learning tasks. The best network survived unchanged to the next generation with its connection weights reset to the initial values it had before training. For all offspring, a copy of the parent's initial weights was then modified by adding a random normally-distributed number with a mean of 0 and a standard deviation of 0.05 to each weight (Batali, 1994). The new offspring networks and the best network from the previous generation were then trained, and the cycle repeated for each generation.

During the pre-language stage, the best network was selected based on the performance on the sequential learning task. After introduction of language into the population, the best network was selected based on performance on the linguistic task with respect to the winning grammar. However, at each generation we only considered networks for selection that maintained their earlier evolved sequential learning abilities. For that purpose we defined a threshold value of minimum sequential learning performance that corresponded to the population average at the end of the pre-linguistic period. The pressure towards maintenance of sequential learning abilities was based on the assumption that this ability would still be advantageous after language was present in the population.

*Linguistic Adaptation.* During each generation five different grammars competed for survival. Linguistic adaptation was simulated by choosing the best learnt grammar as the basis for the next generation. The best learnt grammar survived and reproduced, generating 4 offspring. The initial grammar at the moment of language introduction contained all flexible rewrite rules.[4] Language variation was simulated by mutating the grammars slightly by reassigning the head order of each re-write rule with a certain probability.[5] The mutation rate was 1/12 for each rewrite rule, with 1/3 probability for re-assignment of head-first, head-final or flexible head order, respectively. We let language evolve until it stabilized in the population, that is, after the same grammar was selected for 50 consecutive generations. At that point we stopped the simulations and considered the selected grammar as the winning language. The step-by-step algorithm used to simulate linguistic adaptation can be summarized as follows:

Each generation the following algorithm applies:

1.  Let *Grammar*$_{t-1}$ be the best learnt grammar in the previous generation.
2.  Four offspring are produced from *Grammar*$_{t-1}$ applying the mutation rules.
3.  Train and test separate SRNs on sentences generated by *Grammar*$_{t-1}$ and its 4 offspring.
4.  From *Grammar*$_{t-1}$ and its 4 offspring choose the best learnt grammar, and call it *Grammar*$_t$.
5.  If *Grammar*$_t$ satisfies *Grammar*$_t$ = *Grammar*$_{t-1}$ = *Grammar*$_{t+2}$ = … = *Grammar*$_{t-50}$, then stop the simulation and call *Grammar*$_t$ the winning language, otherwise go to 1.

The results presented here are averaged across 5 different sets of the simulations.

## 3.2 Results and discussion

After the initial 500 generations of training on the sequential learning task alone, the average network performance in a generation had improved significantly ($t(8)$ = 8.51, $p < .0001$) over the performance of the first generation of networks (see Fig. 4). These results are consistent with previous studies (Batali, 1994) in that they demonstrate that it is possible to obtain the Baldwin effect using SRNs trained on complex sequential learning tasks.

After language introduction, networks and languages evolved during many generations before reaching a stable grammar (mean: 110 generations; SD: 36). In all simulations, we found that the same grammar was selected, corresponding to a SOV language. The results are in accord with previous computational work. For example, Van Everbroeck (1999) found that subject-first languages, which make up the majority of language types across the world, were the easier to learn by
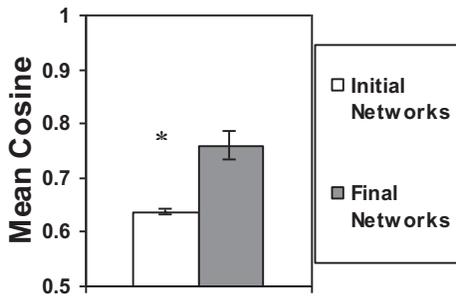
**Figure 4.** Comparison of average performance of the initial networks (white) and final networks (dark) after 500 generations of training in the sequential learning task

recurrent networks (a variation on the SRN). Moreover, these findings are consistent with previous results (e.g., Kirby, 1998), in that the head order of the winning grammar was highly consistent: Five out of six rewrite rules had a head-first order, while head final order was only selected for the VP-rule. Interestingly, in all simulations flexible rewrite rules tended to disappear while consistency tended to increase over time (see Fig. 5). This trend highlights the role of cultural transmission in the emergence of head-order consistency as a result of learning-based constraints.

We found that linguistic adaptation produced a significant improvement in language-learning performance while biological adaptation produced no measurable effect. In order to quantify biological adaptation, we compared the average performance of the initial and final population (networks) when trained on the same language (winning grammar). As illustrated in Fig. 6a, biological adaptation produced no significant improvement in population performance ($t(8) = 0.82$, $p < .43$).
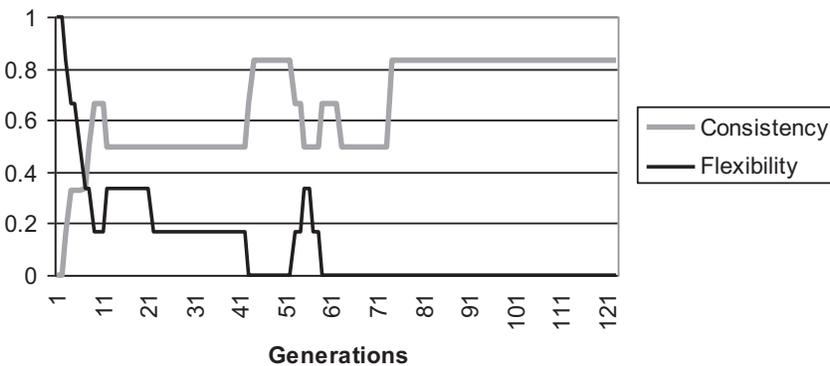


**Figure 5.** Evolution of the rewrite rules' consistency and flexibility over time. Consistency is defined as the proportion of rewrite rules that share the same head order. Flexibility is defined as the proportion of flexible rewrite rules.

In order to measure the effect of linguistic adaptation, we trained the same population (the final generation of networks) on different grammars. When the networks were trained on the winning language, the average population performance was significantly better than when they were trained on the initial grammar ($t(8) = 19.73$, $p < .0001$) (Fig. 6b). As an additional measure of the effect, we compared the average performance of the population when the networks were trained on five random grammars and the winning grammar. The average performance of the networks trained on the winning grammar was significantly better than the average performance of the networks trained on random grammars ($t(8) = 17.49$, $p < .0001$).
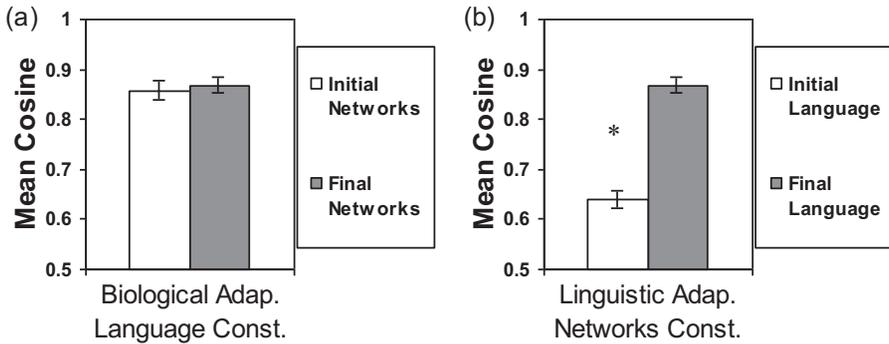


**Figure 6.** a) Comparison of initial and final network performance tested on a fixed language (winning grammar); b) Comparison of initial and final language performance, while keeping the network constant (final network).

## 4. Simulation 2: The role of sequential learning constraints

Simulation 1 shows that biological adaptation is ineffective when language and networks co-evolve and there is a pressure toward maintaining sequential learning capacities. However, it is not clear whether the Baldwin effect would be effective in our evolutionary framework in the absence of a pressure to maintain sequential learning performance. Simulation 2 is designed to test this possibility. Crucially, we manipulated the absence/presence of sequential learning constraints. In Simulation 2a, networks were allowed to evolve with no pressure toward maintaining sequential learning capacities. In Simulation 2b, the pressure toward maintaining sequential learning abilities was reinstated. In order to test for the specific role of sequential learning constraints in preventing the Baldwin effect from occurring, linguistic-adaptation factors were held constant. That is, networks were selected based on their performance on the linguistic task, while the grammar was fixed from the first generation.

### 4.1 Method

#### 4.1.1    *Networks*
The networks were the same as those that constituted the population at the moment of language introduction in Simulation 1. All parameters were the same as in Simulation 1.

#### 4.1.2    *Materials*
Networks were trained on the grammar corresponding to the winning language from Simulation 1.

#### 4.1.3    *Procedure*
We allowed the networks to evolve biologically during the same number of generations necessary to reach a stable grammar in Simulation 1. We simulated absence of linguistic adaptation by keeping language constant throughout the simulations. In Simulation 2a, biological adaptation was simulated similarly to Simulation 1, but the networks were selected purely on their linguistic performance and no constraints toward sequential learning were imposed. In Simulation 2b, the pressure toward maintaining sequential learning abilities was reinstated, and biological adaptation was simulated exactly as in Simulation 1. As in Simulation 1, the results are averaged across five different sets of simulations.

### 4.2 Simulation 2a: Pure biological adaptation

#### 4.2.1    *Results and discussion*
The networks' average performance on the linguistic task increased significantly over time ($t(8) = 5.47$, $p < .001$) (Fig. 7a), showing that it is possible to obtain effective biological adaptation under these conditions.

Simulation 2a differs from Simulation 1 In two fundamental aspects: First, the pressure to maintain sequential learning abilities is absent, and, second, language is constant across generations. Recently, Chater, Reali & Christiansen (2009) conducted simulations suggesting that genes for universal grammar could only co-evolve with aspects of language that are stable in the linguistic environment. They argue that language is a "moving target" over time, and therefore, it would not provide a stable environment for biological assimilation to take place. Thus, it could be possible that the inefficacy of biological adaptation in Simulation 1 is due to the presence of language change and not the sequential learning constraints. In Simulation 2b, the pressure toward maintaining sequential learning capacities was reinstated similar to Simulation 1, but the language was kept constant. Thus, this design provides a way to test the specific contribution of sequential learning
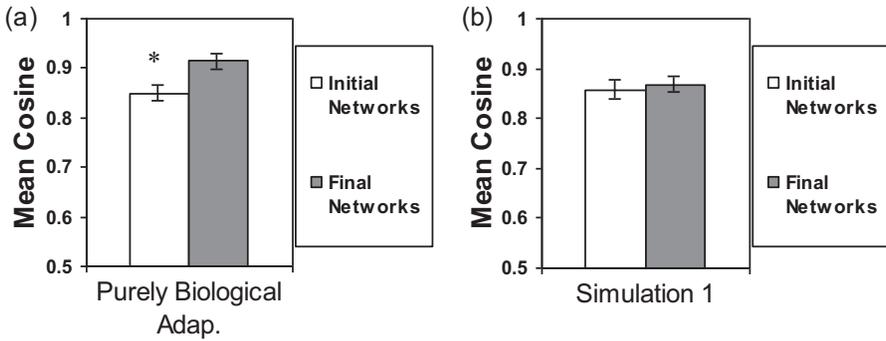
**Figure 7.** a) Results from Simulation 2a: Performance of initial and final networks trained on a fixed language (winning language in Simulation 1) when no pressure toward sequential learning was imposed; b) Comparison of initial and final network performance in Simulation 1 (tested on the winning language) when implementing pressure toward maintaining the evolved sequential abilities.

constraints in preventing biological assimilation of linguistic-specific knowledge in our evolutionary framework.

### 4.3 Simulation 2b: The role of sequential learning constraints

#### 4.3.1 *Results and discussion*

Similarly to Simulation 1, we found that the evolved networks were not significantly better than the initial ones ($t(8) = 1.41$, $p < .195$) (Fig. 8a), indicating that the pressure toward maintaining sequential learning abilities played a causal role in preventing biological adaptation in Simulation 1 (Fig. 8b).
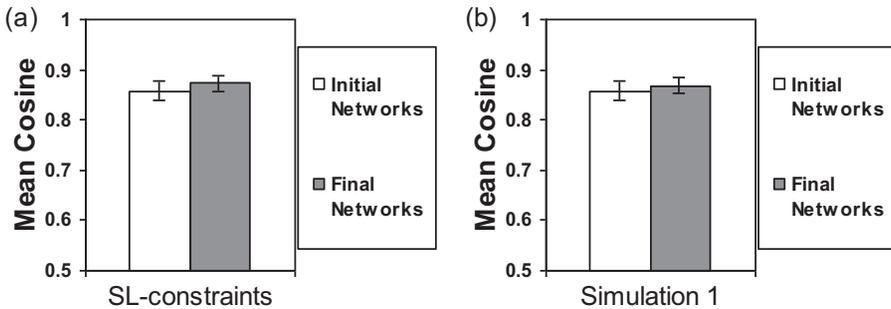


**Figure 8.** a) Results from Simulation 2b: Performance of initial and final networks trained on a fixed language (winning language in Simulation 1) when mainlining a pressure toward sequential learning; b) Comparison of initial and final network performance in Simulation 1 (tested on the winning language) when implementing pressure toward maintaining the evolved sequential abilities.

Overall, the results suggest that gradual assimilation of linguistic knowledge may not be possible when the underlying neural machinery has to accommodate other sequential learning tasks. However, there is a possible caveat to this conclusion:[6] During the initial stage, biological adaptation to the sequential learning task places the networks in a particular part of the evolutionary search space before language is introduced. During the second stage, further evolution is influenced by the continued presence or absence of the sequential learning task. Thus, it is possible that during the initial stage of the simulation, the network weights were moved into a local optimum from which they cannot escape when the sequential learning task is still part of the fitness function after language is introduced. Another possibility is that, irrespective of the starting-point for evolution, a set of initial weights cannot be modified to improve performance on both the sequential task and the linguistic task together. To determine which of these may be the case we ran a control version of Simulation 1 in which the initial stage of adaptation to sequential learning was removed. At the beginning of Stage 2, the initial set of weights was randomized and the performance of the networks on the sequential learning task was measured. The observed (baseline) performance on the sequential learning task was used to establish the fitness function: Networks were selected based on their linguistic performance provided that they maintained the (non-evolved) baseline performance on sequential learning. We let networks and grammars evolve as defined in Stage 2 of the original simulation. As before, the simulation was stopped when the same grammar was selected for 50 consecutive generations. The results were qualitatively the same as before: When a pressure toward *not getting worse* at the sequential learning task was imposed, the Baldwin effect failed to occur ($t(8)=0.79$; $p=0.44$). This suggests that, irrespectively of the set of initial weights, the existence of a pressure toward maintaining performance on the sequential learning task prevents the occurrence of biological adaptation for language.

## 5. General discussion

Fueled by theoretical constraints derived from recent advances in the brain and cognitive sciences, computational modeling has become the paradigm of choice for exploring different theories of language evolution. Even though the use of computer simulations often involves a number of simplifications and abstractions, the advantage of this approach is that specific constraints and/or interactions between constraints can be studied under controlled circumstances. In our case, the simplifications of the linguistic and sequential-learning tasks are at par with many existing models of these types of cognitive behaviors in psychology

and cognitive science (e.g., Boyer, Destrebecqz & Cleeremans, 1998; Christiansen & Chater, 1999; Elman, 1990, 1993; Servan-Schreiber, Cleeremans & McClelland, 1991). Perhaps more importantly, recent work has indicated that such SRN models can be scaled up to deal with more natural sequential-learning tasks (Botvinick & Plaut, 2004) and full-blown corpora of child-directed speech (Reali, Christiansen & Monaghan, 2003).

Together, the simulation results cast doubts on the Baldwin effect as a potential explanation for how a putative universal grammar could have evolved by Darwinian means. But how can we then explain the existence of linguistic universals? An answer may be found in Simulation 1, demonstrating how cultural transmission can help explain linguistic universals such as head-order consistency. Importantly, our simulations go beyond previous work invoking cultural transmission-based explanations of consistent head ordering (e.g., Kirby, 1998). Given that the task of the networks was to predict the grammatical roles of the incoming words – that is, who did what to whom – linguistic adaptation in our simulations not only resulted in the emergence of a more structurally consistent language, but also a language that is easier to interpret. The results add to a growing bulk of work suggesting that some apparently arbitrary aspects of linguistic structure may be functional in terms of learning and processing limitations (e.g., Ellefson & Christiansen, 2000; Kirby, 1998; 1999; O'Grady, 2005; Smith, Brighton & Kirby, 2003; Van Everbroeck, 1999). For example, Smith et al. (2003) used modeling techniques to show how compositional structure in language might have resulted from the complex interaction of learning constraints and cultural transmission. O'Grady (2005) has recently proposed that apparent idiosyncratic binding constraints governing pronominal reference may result from pragmatic factors during processing. In a different series of studies, it has been suggested that subjacency constraints may arise from cognitive constraints on sequential learning (Ellefson & Christiansen, 2000). Moreover, using rule-based language induction, Kirby (1999) accounted for the emergence of typological universals as a result of domain-general learning and processing constraints (see Christiansen & Chater, 2008, for a review).

Simulation 1 showed that when language and learners were allowed to co-evolve, no biological assimilation occurred if networks were required to maintain the same level of performance on sequential learning as obtained before language was introduced into the population. These findings are consistent with recent studies challenging the plausibility of biological assimilation of linguistic knowledge (Chater, Reali & Christiansen, 2009; Christiansen, Reali & Chater, 2006; Kirby & Hurford, 1997; Munroe & Cangelosi, 2002; Yamauchi, 2001). For example, Christiansen, Reali and Chater (2006) conducted a series of computational studies to investigate the circumstances under which universal linguistic constraints might get genetically fixed in a population of language learning agents. The results indicated

that under assumptions of linguistic change, only functional, but not arbitrary, features of language can become genetically fixed. The simulations presented herein illustrate yet another problem with the adaptationist view: The gradual assimilation of linguistic knowledge may not be possible when the underlying neural machinery has to accommodate other sequential learning tasks.

Neural network models trained on corpora encoded in the form of lexical categories are widely used in computational linguistics. However, some caveats to the representational scheme used in our simulations should be noted. For example, it is clear that learners are not provided directly with such "tagged" input. Rather, they have to bootstrap both lexical categories and syntactic constraints concurrently. One way of doing this may involve the combination of distributional information with other kinds of cues during language learning (e.g., Monaghan, Christiansen & Chater, 2007). Moreover, some aspects of natural languages – such as the mapping between form and meaning – are not captured in the input/output representation used in the present simulations. Most connectionist models are restricted to model syntactic aspects of language. However, they are based on the assumption that purely distributional aspects of language are closely entwined with language meaning. Along these lines, natural language processing is viewed as an attempt to retrieve meaning from linguistic form (see Elman, 1991 for further discussion).

The SRN incorporates certain important biases on the learning of sequential structure (Christiansen & Chater, 1999). The importance of exploring such endogenous inductive biases has been recently demonstrated by the work of Griffiths and Kalish (2007) and Kirby, Dowman and Griffiths (2007). Using learning algorithms based on the principles of Bayesian inference, Grifiths and Kalish studied the consequences of iterated learning. In their simulations, Bayesian learners combine prior inductive biases with the evidence provided by linguistic data to compute a posterior distribution over all possible languages. They found that iterated learning converges to a distribution over languages that is determined by the learner's prior inductive biases. These results indicate that learning biases have a strong influence on linguistic adaptation. Recently, Kirby, Dowman and Griffiths (2007) used similar methods to show that when learners select languages with maximum posterior probability, the final distribution over languages is *also* determined by factors of cultural transmission, such as the amount of information transmitted between generations. They concluded that, under some learning assumptions, cultural transmission factors can magnify weak endogenous biases.

A crucial assumption adopted here is that language learning and processing shares mechanisms with sequential learning in other domains. A growing number of neuroimaging studies now provide empirical support for this notion (Koelsch et al., 2002; Maess et al., 2001; Patel, 2003, Patel et al., 1998; Friederici et al., 2002;

Petersson et al., 2004). Moreover, recent studies suggested that breakdown of language capacities is associated with impaired sequential learning in non-linguistic tasks (Christiansen et al., 2007; Hoen et al., 2003; Hsu, Christiansen, Tomblin, Zhang & Gómez, 2006; Plante, Gómez & Gerken, 2002). For example, Christiansen et al. (2007) found that agrammatic aphasics, who typically have damage in or around Broca's area, showed decreased performance on a sequential learning task. In a different study, Hsu et al. (2006) showed that specific language impairment is associated with impaired sequential learning. Moreover, Hoen et al. (2003) found that increased performance on a visual sequence-learning task in agrammatic aphasics resulted in improvements in their abilities to understand certain complex linguistic constructions. Thus, from an evolutionary perspective, it seems reasonable to assume that language originally emerged based on pre-existing learning and processing mechanisms (e.g., Kirby & Christiansen, 2003; Pinker & Bloom, 1990). However, if language originally emerged by piggybacking on prior sequential-learning mechanisms, it is unlikely that language could have "taken over" these mechanisms because being able to extract and process sequential information would still have been crucial for negotiating the social and physical environment of the hominids.

A further assumption of our simulations is that there have been specific biological adaptations for better sequential learning abilities in the hominid lineage. Recent work in human molecular genetics and comparative genomics relating to the *FOXP2* gene suggests that a genetic adaptation for this type domain-general learning may indeed have taken place in recent human evolution (Fisher, 2006). Mutations to the *FOXP2* gene result in severe speech and orofacial motor impairments (Lai et al., 2001; MacDermot et al., 2005). Studies of *FOXP2* expression in mice and imaging studies of an extended family pedigree with *FOXP2* mutations have provided evidence that this gene is important to the development and function of the corticostriatal system as well as other neural systems (Lai et al., 2003). These systems have been shown in other studies to be important for sequential and other types of procedural learning (Packard & Knowlton, 2002). In family members affected by FOXP2 mutations, the volume of the caudate was found to be smaller than for unaffected family members (Watkins et al., 2002). Crucially, preliminary findings from a mother and daughter with a translocation involving *FOXP2* indicate that they have problems with both language and sequential learning (Tomblin et al., 2004). Cross-species comparisons have shown that *FOXP2* is highly conserved across species, showing evidence of only 3 amino acid changes in the FOXP2 protein since the last common ancestor for mice and humans, some 170 million years ago (Enard et al., 2002). However, two of these changes happened after the split between humans and chimps about 5-6 millions ago, and statistical analyses suggest that these changes happened rapidly and got fixed in the human

population about 200,000 years ago. Thus, the current knowledge regarding the *FOXP2* gene is consistent with the kind of evolutionary scenario detailed in our simulations, but not, as previously thought, with the evolution of some aspects of universal grammar.

Finally, we note that in our simulations, we have approximated biological adaptation by selecting the best-learning network's initial connection weights at each generation. Therefore, the simulation results pertain to a gradual assimilation of innate knowledge encoded in fine-grained patterns of connectivity (see also, Batali, 1994; Munroe & Cangelosi, 2002). This conforms to the standard way of characterizing the knowledge of a network in terms of the strength of its connection weights (McClelland, Rumelhart & Hinton, 1986). Elman et al. (1996) have described this definition of innate knowledge as the strongest and most specific form of nativism. Such *representational* nativism would allow for an innately specified encoding of detailed rules of, say, grammar, physics or theory of mind (for discussion see chapter 7, Elman et al., 1996). Although our simulations suggest that linguistic assimilation at the level of representational innateness may not be effective when language evolution also incorporates sequential-learning constraints and linguistic change, they do not address whether the Baldwin effect could potentially occur at the level of *architectural* constraints. These constraints comprise innate specifications of the structural aspects of the networks, including the computational properties of individual units and the general characteristics of layering and connectivity within a specific region of the network. However, changes to such architectural constraints are more likely to be reflected in differences in general learning abilities, rather than the kind of domain-specific linguistic knowledge characteristic of a universal grammar (Deacon, 2003).

In sum, our simulations illustrate the effectiveness of linguistic adaptation to improve language learnability and challenge the plausibility of biological assimilation of linguistic-specific knowledge. Together, the findings indicate that the emergence linguistic structure may have resulted from the complex interaction of domain-general architectural constraints and the process of linguistic adaptation through cultural transmission.

### Acknowledgments

## Notes

**1.** We adopted this input representation in the sequential learning task because the linguistic task required a larger vocabulary and we used the same networks for both tasks.

**2.** We use 'adpositional phrase' to denote that the rewrite rule may involve either a prepositional phrase or a postpositional phrase depending on the head-order; 'possessive phrase' is used to denote rules involving possessive relationships between two nouns either through a possessive marker (such as *'s* in *the general's daughter*) or adpositional constructions (such as the use of *of* in *the daughter of the general*).

**3.** The cosine measure ranges from 0 to 1, with 1 corresponding to perfect performance.

**4.** The initial state of language when introduced can be seen as a lexical-based proto-language with no syntactic constraints imposed apart from the presence of at least a subject noun and a verb. We remain agnostic with regard to the question of the origin of proto-language, but base our simulations on the historical fact that at some time in the human lineage language *did* emerge.

**5.** Our use of grammar mutation to introduce linguistic variation is a computational simplification of the effects of cultural transmission. Theoretically, we envisage that differences in learning and use of language among interacting agents would drive this process.

**6.** We are thankful to an anonymous reviewer for suggesting this possibility.

## References

Baldwin, J.M. (1896). A new factor in evolution. *American Naturalist, 30,* 441-451.

Batali, J. (1994). Innate biases and critical periods: combining evolution and learning in the acquisition of syntax. In R. Brooks, & P. Maes (Eds.), *Artificial Life 4: Proceedings of the Fourth International Workshop on the Synthesis and Simulations of Living Systems* (pp. 160-171). Redwood City, CA: Addison-Wesley.

Batali J. (1998). Computational simulations of the emergence of grammar. In J. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases* (pp. 405-426). New York: Cambridge University Press.

Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, *111*, 395-429.

Boyer M., Destrebecqz A., & Cleeremans A. (1998). The serial reaction time task: Learning without knowing, or knowing without learning? In *Proceeding of the Twentieth Annual Meeting of the Cognitive Science Society* (pp. 167-172). New Jersey: Erlbaum.

Briscoe, E. (2002). Grammatical acquisition and linguistic selection. In E. Briscoe (Ed.), *Linguistic Evolution through language acquisition: Formal and computational models* (pp. 255–300). New York: Cambridge University Press.

Briscoe, E. (2003). Grammatical assimilation. In M.H. Christiansen, & S. Kirby (Eds.), *Language evolution* (pp. 295-316). New York: Oxford University Press.

Bybee, J. (2002). Sequentiality as the basis of constituent structure. In T. Givón, & B. Malle (Eds.), The evolution of language out of pre-language (pp. 107-132). Philadelphia, PA: John Benjamins.

Cangelosi, A. (1999). Modeling the evolution of communication: from stimulus associations to grounded symbolic associations. In D. Floreano, J. D. Nicoud, & F. Mondada (Eds.), *Advances in artificial life (Proceedings ECAL99 European Conference on Artificial Life)* (pp. 654-663). Berlin: Springer-Verlag.

Chater, N., Reali, F. & Christiansen, M.H. (2009). *Restrictions on biological adaptation in language evolution*. Proceedings of the National Academy of Sciences, 106, 1015–1020.

Christiansen, M.H. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure*. Unpublished PhD dissertation. University of Edinburgh, Scotland.

Christiansen, M.H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23,* 157-205.

Christiansen, M.H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral & Brain Sciences*, 31, 487–558.

Christiansen, M.H., Conway, C.M., & Onnis, L. (2007). Neural responses to structural incongruencies in language and statistical learning point to similar underlying mechanisms. In D.S. McNamara & J.G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 173-178). Austin, TX: Cognitive Science Society.

Christiansen, M.H., Dale, R., Ellefson, M.R., & Conway, C.M. (2002). The role of sequential learning in language evolution: computational and experimental studies. In A. Cangelosi, & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 165-187). London: Springer-Verlag.

Christiansen M.H., & Devlin J.T. (1997). Recursive inconsistencies are hard to learn: A connectionist perspective on universal word order correlations. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 113-118). Mahwah, NJ: Lawrence Erlbaum Associates.

Christiansen, M.H., Kelly, L., Shillcock, R. & Greenfield, K. (2007). *Impaired artificial grammar learning in agrammatism*. Manuscript under revision.

Christiansen, M.H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences, 7,* 300-307.

Christiansen, M.H., Reali, F. & Chater, N. (2006). The Baldwin effect works for functional, but not arbitrary, features of language. In A. Cangelosi, A. Smith & K. Smith (Eds.), *Proceedings of the Sixth International Conference on the Evolution of Language* (pp. 27-34). London: World Scientific Publishing.

Conway, C.M., & Christiansen, M.H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences, 56*(5), 539–546.

Davidson, I. (2003). The archaeological evidence of language origins: States of art. In M.H. Christiansen, & S. Kirby (Eds.), *Language evolution* (pp. 140-157). New York: Oxford University Press.

Deacon, T. (1997). *The symbolic species: The coevolution of language and the brain*. New York, NY: Norton.

Deacon, T. (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 Baldwin effect reconsidered* (pp. 81-106). Cambridge, MA: MIT Press.

Dunbar, R.I.M. (2003). The origin and subsequent evolution of language. In M.H. Christiansen, & S. Kirby (Eds.), *Language evolution* (pp. 219-234). New York: Oxford University Press.

Ellefson, M.R. & Christiansen, M.H. (2000). 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 Society* (pp. 645-650). Mahwah, NJ: Lawrence Erlbaum.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14: 179-211.

Elman, J.L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7*, 195-224.

Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. C*ognition*, *48*, 71–99.

Elman, J. L., Bates E. A., Johnson, M. H., Karmiloff –Smith A., Parisi D., & Plunkett, K. (1996). *Rethinking innateness*. Cambridge, MA: MIT Press.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, *418*, 869–872.

Fisher, S.E. (2006). Tangled webs: Tracing the connections between genes and cognition. *Cognition, 101*, 270-297.

Friederici, A.D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing. *Proceedings of the National Academy of Sciences of United States of America, 99,* 529-534.

Givón, T. (1998). On the co-evolution of language, mind and brain. *Evolution of communication, 2*, 45–116.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science,* 31, 441–480.

Heine, B., & Kuteva, T. (2002). On the evolution of grammatical forms. In A. Wray (Ed.), *Transitions to language* (pp. 376–397), Oxford: Oxford University Press

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

Hoen M., Golembiowski, M., Guyot E., Deprez V., Caplan D., & Dominey, P.F. (2003). Training with cognitive sequences improves syntactic comprehension in agrammatic aphasics. *NeuroReport, 14,* 495-499.

Hurford, J. (1989) Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, 77, 187-222.

Hurford, J., & Kirby, S. (1999). Co-evolution of language size and the critical period. In David Birdsong (Ed.) *Second Language Acquisition and the Critical Period Hypothesis* (pp.39-63), Lawrence Erlbaum.

Jackendoff , R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Kirby, S. (1998). Fitness and the selective adaptation of language. In J.R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 359-383). New York: Cambridge University Press.

Kirby, S. (1999). *Function, selection and innateness: The emergence of language    universals*. Oxford: Oxford University Press.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation* 5(2): 102-110.

Kirby, S. (2002). Natural language from artificial life. *Artificial Life, 8*, 185–215.

Kirby, S., & Christiansen, M.H. (2003). From language learning to language evolution. In M.H. Christiansen, & S. Kirby (Eds.), *Language evolution* (pp. 272-294). New York: Oxford University Press.

Kirby, S., Dowman, M., & Griffiths, T. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences, 104,* 5241-5245.

Kirby, S., & Hurford, J. (1997) Learning, culture and evolution in the origin of linguistic constraints. In P. Husbands, & I. Harvey (Eds.), *Fourth European conference on artificial life* (pp. 493-502). Cambridge: MIT Press.

Kirby, S., & Hurford J. (2002). The emergence of linguistic structure: an overview of the iterated learning model. In A. Cangelosi, & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121-147). London: Springer-Verlag.

Knasnicka, V., & Pospichal, J. (1999). An emergence of coordinated communication in populations of agents. *Artificial Life, 5,* 318-342.

Lai, C. S.L., Fisher, S.E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413,* 519-523.

Lai, C. S. L., Gerrelli, D., Monaco, A. P., Fisher, S. E., & Copp, A. J. (2003). FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain, 126,* 2455–2462.

Lee, Y.S. (1997). Learning and awareness in the serial reaction time. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society.* (pp. 119-124). Hillsdale, NJ: Lawrence Erlbaum Associates.

Levinson, S.C (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Livingstone, D., & Fyfe, C. (2000). Modelling language-physiology coevolution. In C. Knight, M., Studdert-Kennedy and J. R. Hurford (Eds.), *The emergence of language: Social function and the origins of linguistic form* (pp. 199- 215). Cambridge University Press.

MacDermot, K. D., Bonora, E., Sykes, N., Coupe, A. M., Lai, C. S. L., Vernes, S. C., et al. (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *American Journal of Human Genetics, 76*, 1074–1080.

Maess, B., Koelsch, S., Gunter, T., & Friederici, A.D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience, 4*: 540-545.

McClelland, J.L., Rumelhart, D.E., & Hinton, G.E. (1986). The appeal of parallel distributed processing. In J.L McClelland, & D.E. Rumelhart (Eds.), *Parallel distributed processing, Vol. 1* (pp. 3-44). Cambridge, MA: MIT Press.

Monaghan, P., Christiansen, M.H. & Chater, N. (2007). The Phonological-Distributional Coherence Hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology,* 55, 259–305.

Munroe S., & Cangelosi A. (2002). Learning and the evolution of language: the role of cultural variation and learning cost in the Baldwin Effect. *Artificial Life, 8*, 311-339

Nowak, M. A., Komarova, N. L., & Nyogi, P. (2002). Computational and evolutionary aspects of language. *Nature, 417*: 611–17.

O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Mahwah, NJ: Erlbaum.

Packard, M. & Knowlton, B. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience, 25,* 563-593.

Patel, A.D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6,* 674-681.

Patel, A.D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P.J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*: 717-733.

Petersson, K.M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science, 28*, 383-407.

Pinker, S. (1994). *The language instinct*. New York: HarperCollins/Morrow.

Pinker, S. (2003). Language as an adaptation to the cognitive niche. In M.H. Christiansen, & S. Kirby (Eds.), *Language evolution* (pp. 16-37). New York: Oxford University Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13*, 707–784.

Ragir, S. (2002). Constraints on communities with indigenous sign languages: clues to the dynamics of language origins. In A. Wray (Ed.), *Transitions to language* (pp. 272-296). Oxford: Oxford University Press.

Reali, F., Christiansen, M.H. & Monaghan, P. (2003). Phonological and distributional cues in syntax acquisition: Scaling up the connectionist approach to multiple-cue integration. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 970-975). Mahwah, NJ: Lawrence Erlbaum.

Servan-Schreiber, D., Cleeremans, A., & McClelland, J.L. (1991). Graded state machines: The representation of temporal dependencies in simple recurrent networks. *Machine Learning*, *7*, 161-193.

Smith, K. (2004). The evolution of vocabulary. *Journal of Theoretical Biology, 228 (1)*, 127-142.

Smith, K. (2002). Natural selection and cultural selection in the evolution of communication. *Adaptive Behavior, 10 (1)*, 25-44.

Smith, K., Brighton, H., & Kirby S. (2003). Complex systems in language evolution: the cultural emergence of compositional structure. *Advances in Complex Systems, 6,* 537-558.

Tomasello, M. (2003). On the different origins of symbols and grammar. In M.H. Christiansen, & S. Kirby (Eds.), *Language evolution* (pp. 94-110). New York: Oxford University Press.

Tomblin, J.B. Shriberg, L. Murray, J., Patil, S., & Williams, C. (2004) Speech and language characteristics associated with a 7/13 translocation involving FOXP2. *American Journal of Medical Genetics*, *130B*, 97.

Van Everbroeck, E. (1999). Language type frequency and learnability: A connectionist appraisal. In *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 755-760). Mahwah, NJ: Lawrence Erlbaum Associates.

Waddington, C.H. (1940). *Organisers and genes.* Cambridge: Cambridge University Press.

Watkins, K. E., Vargha-Khadem, F., Ashburner, J., Passingham, R., Connelly, A., Friston, K. J. et al. (2002). MRI analysis of an inherited speech and language disorder: structural brain abnormalities. *Brain, 125,* 465-478.

Weber, B.H., & Depew, D.J. (Eds.) (2003). *Evolution and learning: The Baldwin effect reconsidered.* Cambridge, MA: MIT Press.

Yamauchi, H. (2001). The difficulty of the Baldwinian account of linguistic innateness. In J. Kelemen and P. Sosík (Eds.), *ECAL01* (pp. 391-400). Prague: Springer.

## *About the authors*

**Morten H. Christiansen** received his PhD in Cognitive Science from the University of Edinburgh in 1995. He is an Associate Professor in the Department of Psychology and Co-Director of the Cognitive Science Program at Cornell University as well as an External Professor at the Santa Fe Institute. His research focuses on the interaction of biological and environmental constraints in the processing, acquisition and evolution of language, which he approaches using a

variety of methodologies, including computational modeling, corpus analyses, psycholinguistic experimentation, neurophysiological recordings, and molecular genetics.

**Florencia Reali** obtained her M.S. in Biological Sciences in 2002 from Universidad de la República, Montevideo, Uruguay. She entered graduate studies at Cornell University where she worked on computational modeling of cognitive processes under the supervision of Morten H. Christiansen. After receiving her PhD in Psychology in 2007, she became a postdoctoral fellow in Thomas L. Griffiths' lab at UC Berkeley. Her research combines behavioral experiments and probabilistic models to study various aspects of language learning and processing. Her current interests include the exploration of some theoretical aspects of language evolution, including the interaction between cultural transmission, biological adaptation and individual learning.

*Author's address*

Florencia Reali
Institute of Cognitive and Brain Sciences,
UC Berkeley
Berkeley, CA 94720
USA

florencia.reali@gmail.com