

## 10 Genetic Biases in Language: Computer Models and Experimental Approaches

---

*Rick Janssen and Dan Dediu*

### **Abstract**

Computer models of cultural evolution have shown language properties emerging on interacting agents with a brain that lacks dedicated, nativist language modules. Notably, models using Bayesian agents provide a precise specification of (extra-)linguistic factors (e.g., genetic) that shape language through iterated learning (biases on language), and demonstrate that weak biases get expressed more strongly over time (bias amplification). Other models attempt to lessen assumption on agents' innate predispositions even more, and emphasize self-organization within agents, highlighting glossogenesis (the development of language from a nonlinguistic state). Ultimately however, one also has to recognize that biology and culture are strongly interacting, forming a coevolving system. As such, computer models show that agents might (biologically) evolve to a state predisposed to language adaptability, where (culturally) stable language features might get assimilated into the genome via Baldwinian niche construction. In summary, while many questions about language evolution remain unanswered, it is clear that it is not to be completely understood from a purely biological, cognitivist perspective. Language should be regarded as (partially) emerging on the social interactions between large populations of speakers. In this context, agent models provide a sound approach to investigate the complex dynamics of genetic biasing on language and speech.

### **10.1 Introduction**

#### *10.1.1 Biasing Language*

In this chapter, we argue not only that the best approach to understanding the origins and present-day diversity of language is rooted in evolutionary

theory, but also that extra-linguistic factors, more specifically biological ones in our genes, may play an important role in shaping language. Likewise, these factors do not act in a void, but interact with multiple constraints and affordances on different scales in parallel. So-called *cultural evolution of language* (Section 10.1.2) must thus be seen in a rich context (partially) molded by the biological and cognitive entities that ultimately acquire, use, and transmit language – us. Important factors in this context are therefore represented not only by the brain – it has been recognized for a while now that the brain indeed shapes language (Christiansen & Chater 2008) – but also by the anatomy and physiology of the vocal tract and hearing organs. Just to illustrate, it has been recently suggested (Butcher 2006) that the very high rates of chronic otitis media (an infection of the middle ear that impacts hearing) affecting Australian Aboriginal children might explain striking features of the phonological systems of the Australian languages such as a lack of fricatives and many distinctive places of articulation. This process of *biasing*, whereby extra-linguistic factors, ultimately with a genetic basis, can affect cultural evolution of language, has been suggested to be a rather general influence playing a role in explaining not only universal tendencies (when these biases are shared across the whole human species) but also linguistic diversity (when the biases differ between human populations in magnitude or direction) (Ladd, Dediu & Kinsella 2008; Dediu 2011).

There are multiple lines of evidence supporting genetic biasing of language and many interesting directions to explore,<sup>1</sup> but we focus on a very specific set of questions: What can we conclude about the nature and effects of such biases from the body of computational modeling work and experimental approaches (using both human subjects as well as animal models) on language change and evolution? To this end, we begin by discussing some fundamental notions necessary for a cultural evolutionary approach and the influence of genetic biases (Section 10.1.2), followed by an overview of some relevant computer models (such as various iterated learning approaches [Sections 10.2.1 and 10.2.2] but also models that do not belong to this tradition [Section 10.2.4]). Computational models such as these are particularly interesting to investigate processes that develop on long time scales, since they allow for experimental manipulation not available otherwise (de Boer & Fitch 2010). However, a sound empirical foundation, possible in series with other models, is hereby essential in order to make testable predictions, such as those corroborated by the experimental results in Section 10.2.3. Finally, we discuss models that address the possibility

<sup>1</sup> Both authors are part (together with Dr. Scott Moisiuk) of the **Genetic Biases in Languages and Speech (G3bils)** project, which actively investigates the influence of the vocal tract as a biasing factor on phonetics and phonology (<http://www.mpi.nl/departments/language-and-genetics/projects/genetic-biasing-in-language-and-speech>).

of feedback from culture into the genome through the Baldwin Effect (Section 10.3). This overview of a diverse literature (and we must highlight the fact that few studies were designed with genetics in mind) suggests that while genetic biases can indeed affect language, it is still too early to draw any general conclusions regarding the strength required for such biases to become manifest and be measurable in human populations (Section 10.4).

### *10.1.2 Cultural Evolution of Language*

Many accounts of the nature, origins, and evolution of language do not consider evolutionary processes to play any important role (e.g., Chomsky 1986). Even the field of historical linguistics (e.g., Campbell & Poser 2008), which tries to understand the factors, processes, and outcomes of language change across time, does not have an evolutionary outlook and seems generally rather critical of approaches that use such concepts and methods (see, for example, the quite cold reception of modern phylogenetic approaches to questions of language relatedness, the dating of proto-languages, and the expansion of language families such as Indo-European and Austronesian (e.g., Gray & Atkinson 2003; Pagel, Atkinson & Meade 2007; Dunn, Greenhill, Levinson & Gray 2011; Bouckaert, Lemey, Dunn, Greenhill, Alekseyenko, Drummond, Gray, Suchard & Atkinson 2012)). On the other hand, there are other proposals that consider biological evolution to be the main explanatory factor behind the human use of language (e.g., Pinker & Bloom 1990), but they miss the intervening causal role played by cultural evolution by emphasizing biological nativism.

The emphasis on biological evolution is not surprising. Darwinian theory, based on the principles of replication, variation, and selection, has proven to be an immensely powerful approach to explain biological complexity (Carroll 2005). In a nutshell (and glossing over many fascinating aspects of evolutionary biology), when a population of organisms reproduces, slight variations will be introduced in the offsprings' genome by mutations. These mutations are essentially random, most of them having a neutral or negative effect on an organism's phenotype. Mutations might, for instance, cause inheritable diseases such as sickle-cell anemia (OMIM<sup>2</sup> 603903) or developmental speech dyspraxia (OMIM 602081). A small number of mutations, however, might have positive effects. They might, for example, give an animal a slightly increased resistance to certain pathogens, enlarged cardiovascular capacity, or enhanced cognitive capabilities. If these improvements are small, just one of them is of course unlikely to be saliently noticeable. However, advantageous mutations

<sup>2</sup> For the sake of brevity, we will refer to OMIM (Online Mendelian Inheritance in Man; <http://omim.org>) unique identifiers that give access to brief up-to-date descriptions and the relevant literature.

will accumulate as an effect of selection whereby the organisms with an increased fitness (in part ascribable to these advantageous mutations) are more likely to reproduce, transmitting the mutations to their offspring.

A common misconception is that evolution is teleological (Hanke 2004). For instance, the notion that we, *Homo sapiens*, have evolved to have large cognitive capabilities is most likely a matter of circumstances more than anything else.<sup>3</sup> As a matter of fact, natural selection should be viewed as a system merely acting as a filter on the existing variation in a population, in many ways similar to many optimization algorithms used in computer science (e.g., metaheuristics). Thus, evolution produces ad hoc solutions that are merely appropriate for the situation at hand, without any notion of design, aesthetics, elegance, rational insight, or intentionality. This lack of foresight is strikingly apparent when looking at what could be considered “design errors” such as photoreceptors pointing away from the direction light strikes the retina in vertebrates.<sup>4</sup> Design errors demonstrate that showing that Darwinian processes are at work does not warrant the conclusion that they should necessarily result in sophistication and, conversely, that if we observe sophistication, Darwinian processes are per definition responsible.<sup>5</sup> Nevertheless, natural selection is widely considered the most powerful explanation for the origin of biological complexity on phylogenetic time scales (i.e., pertaining to the formation of taxonomic groups). As such, the proposition that biological evolution is responsible for the complex system of language as well seems to be a logical one (e.g., Pinker & Bloom 1990). Upon closer consideration, however, given the fact that languages change much faster than any biological evolution to fully account for it (Christiansen & Chater 2008), this appears unlikely.

Even though the principles of replication, variation, and selection are simple, biologists have been vigorously debating the “level” they act on. Intuitively, this level might appear to be located at the scale of individuals, i.e., organisms competing with each other for food and mates. The idea of group-level selection on the other hand has been invoked to explain some widely observed behaviors such as altruism as demonstrated by, for instance, the use of distress calls in

<sup>3</sup> There is however considerable debate on the evolution of the human brain, ranging from an effect of social (Dávid-Barrett & Dunbar 2013), sexual (Miller 2001), environmental (Calvin 2002), or other selective pressures.

<sup>4</sup> There is some debate on whether the inverted vertebrate retina is the result of a historically frozen maladaptation or a trade-off between optical and other physiological (e.g., metabolic) costs (Kröger & Biehlaier 2009).

<sup>5</sup> What is described here is also known as the logical fallacy of “affirming the consequent” Although natural selection is one mechanism that explains complexity, other mechanisms might do so comparably well. For instance, ice-crystal or spiral-galaxy formation does not require descent with modification but self-organizes following mere physical, in situ interactions (e.g., Lin & Shu 1964).

groups of meerkats when spotting a predator (Wynne-Edwards 1962; Wynne-Edwards 1986).<sup>6</sup> More recently, the gene-centered view of selection has been popularized, providing an alternative for explaining altruism and introducing the concept of the *extended phenotype* that reaches beyond the confines of the biological organism and includes “expressions” such as bird nests and termite mounds (Hamilton 1963; Williams 1966; Dawkins 1976).

Essentially, the debate on the level of selection centers on the conceptualization of the replicating unit (or *replicator*) that drives evolution. Interestingly, if Darwinism is not confined to the domain of biology alone (and why should it?), these replicators might not just exist at different levels of a biological system (multilevel selection; Okasha 2006; Wilson & Wilson 2008) but also in different domains altogether. One such domain might of course be human language. So, what might a linguistic replicator look like?<sup>7</sup> Consider that a human community engaged in linguistic exchange is producing a population of utterances that are actively used in everyday speech. Similarly to how organisms are “composed” of genes, utterance can be regarded to be composed of *linguemes*, “building blocks” that can change and be recombined to form new utterances (Croft 2000). Examples of these linguemes – linguistic replicators – are phonemes, morphemes, words, or even grammatical rules. While genes are transmitted during reproduction from parent to offspring, linguemes are transmitted from teacher to learner, for instance during, but not restricted to, first language acquisition during childhood. As we know, genes that are likely to be transmitted during reproduction will, *ceteris paribus*, eventually spread throughout a population. Similarly, those linguemes that are likely to be transmitted from teacher to learner will eventually proliferate in a cultural population of linguistic units. Even though there are obvious differences between genes and linguemes, such as the degree of horizontal information flow, the encoding medium, and the speed of change and transmission noise,<sup>8</sup> there are the (striking) common principles of reproduction, variation, and selection (Levinson & Gray 2012) that seem to provide a valid explanation for the complex phenomenon of language without having to rely on biological determinism alone.

<sup>6</sup> The confusion on the level of selection is also visible in the abuse of Darwinian theory in the justification of e.g., eugenics and racialism during the early 20th century.

<sup>7</sup> Even if this question remained unanswered, it would not invalidate the viewpoint that language evolves in a cultural domain. Mendelian genetics has been successfully practiced before the discovery of the encoding medium, DNA. Our current understanding of cultural evolution might be of a comparable advancement.

<sup>8</sup> The term “noise” should be understood in its broadest sense here. The brain has an active, heuristic role in selecting which linguemes are transmitted (Christiansen & Chater 2008). Noise in cultural selection is therefore not as uniform as in genetics but likely structured to some degree (see e.g., Farrell, Wagenmakers & Ratcliff 2006 for the presence of pink noise – structured noise showing self-similarity – in cognitive performance).

Still, we have not discussed a good analogue of selective pressure in biology. Section 10.2 addresses this issue by explaining selection on linguemes as a *transmission bottleneck* by means of a series of computer modeling experiments. Interestingly, many of these models imply, as briefly mentioned in Section 10.1.1, that cultural evolution of language converges to particular states and that this convergence is influenced by biasing factors. Examples of these factors could be environmental (e.g., altitude, humidity) (Everett, Blasi & Roberts 2015), social (e.g., peer pressure, conformity bias, sexual selection), cognitive (e.g., working memory capacity, language impairments like aphasia), genetic (e.g., *FOXP2* (OMIM 605317), *ASPM* (OMIM 605481), *MCPHI* (OMIM 607117)), or anatomical (e.g., hard-palate curvature, jaw length) (Henrich & McElreath 2003). Biases such as these might co-exert a likely very subtle selective pressure in cultural evolution of language, potentially providing an explanation for why languages differ. Importantly, these differences do not imply strong biases per se: It is very unlikely that human languages differ because populations have different innate, all-or-nothing cognitive or anatomical predispositions that allow for or prohibit certain language features. Instead, genetics might, for example, influence a speaker's "effort" associated with producing certain speech sounds or patterns. These efforts might then, over time, lead to a cascading effect via cultural evolution. In such a case, even the weakest bias might get amplified to the point where it could (even) lead to phonemic discriminability (Ladd et al. 2008; Dediu 2011). In fact, this amplification effect has been observed in computer simulations (Section 10.2.2) and, arguably, in animal models (Section 10.2.3).

As a slightly more concrete example of anatomical biasing factors, it is not hard to imagine a particular hard palate shape that, say, eases ingestion and disturbs the production of particular speech gestures while facilitating others. While speculative, similar propositions have been postulated before. For example, Ladefoged (1984) made the observation that while Italian and Yoruba (a Niger-Congo language) have very similar vowel systems, there are nevertheless slight differences, possibly due to a minor anatomical between-group variance (e.g., a slightly larger mouth opening in Yoruba speakers, lowering the second formant in vowels). Importantly, such a very subtle, almost undetectable bias, would not prevent any Italian from speaking Yoruba if being exposed to it from infancy (and the other way around). On a *glossogenetic* level (i.e., at the time scales concerning historical language change), however, these subtle anatomical biases might become saliently expressed on a cultural, population level. Culturally amplified biases such as these could explain the present-day distributions of features seen in human language, such as Yoruba second formant lowering or word order, not as having arrived abruptly by a chance mutation giving rise to "language modules" (Fodor 1983) or "language acquisition devices" (Chomsky 1965), but as (weakly) emerging from the interactions between large

populations of situated speakers. This then effectively describes a *dynamical system* that slowly gravitates toward particular attractors in an articulation-landscape that is (partially) formed by these biases.

To summarize, language can thus be seen as a system evolving in itself, thereby reducing the plausibility of strong biological, even nativist explanations for its structure, diversity, and evolution. However, this does not imply that extra-linguistic factors are of no effect. When we consider the two evolving systems – biological and cultural – in parallel, it is not hard to see how they might be compared to how organisms, such as predator and prey or parasite and host, are *coevolving* (Richerson & Boyd 2008). Likewise, culture and biology might be exerting reciprocal selective pressure on each other, raising the possibility that, on one hand, biology influences (or biases) culture and, on the other, stable cultural features become “inscribed” into the (biological) genome (Section 10.3). Even without addressing the full complexities of coevolution (Section 10.3), we can already be sure that cultural and biological evolution can by no means be considered independent from each other.

## 10.2 Computer Models of Cultural Evolution

### 10.2.1 Iterated Learning

As discussed in Section 10.1.2, cultural evolution of language centers on the idea that linguistic replicators are in competition with each other, similarly to how genes might be in biology. To investigate cultural evolution in computer simulation, Kirby & Hurford (2002) developed the Iterated Learning (IL) agent model. IL is of particular interest to the topic of genetic biasing not only because it explains some features unique to human language (Hockett 1960) without any (a priori) requirement on biological evolution but also because it allows for a precise specification of the nature and strengths of these biases (Section 10.2.2).

Briefly summarized, IL simulates language transmission from teacher to learner. This works by agents conveying “meanings” to each other by transmitting “utterances.” At the time of this writing, there are many incarnations of the procedure. To start of simply, in Kirby & Hurford (2002), only one pair of teacher and learner exists at any time (a population of two), where the old learner replaces the teacher when a new learners is introduced. This population model is now often considered to be a sort of sequential teacher-learner chain (Fig. 10.1). The conversion of meanings to utterances (and vice versa) is determined by the agents’ internal rules (i.e., a “grammar”).

In the chain model, once all conveyed meanings are learned, the learner will take the role of teacher and attempt to convey the interpreted meaning to the next learner. By transmitting meanings from agent to agent, language transmission, and thereby cultural evolution of language, is modeled. The longer

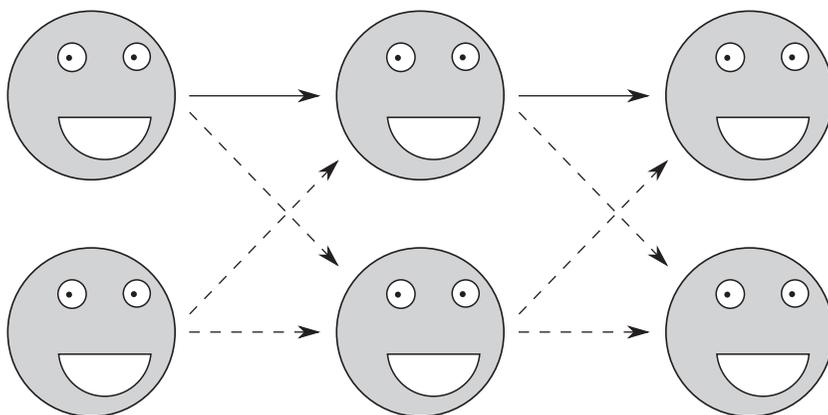


Figure 10.1 The IL social structure. Agents transmit utterances to each other following the arrows. Solid arrows mark those channels available in the monadic experiments, modeling vertical information transfer (e.g., from single parent to single learner). Dashed arrows mark those communication channels available in polyadic setups, modeling oblique flow (e.g., teachers that teach multiple learners, while learners learn from multiple teachers). The classical IL paradigm does not normally allow for horizontal communication (e.g., between agents from the same “generation”). Note that there is no biological evolution in IL: agents’ architecture remains unchanged.

the chain, the more transmissions occur and the larger the time-frame that is modeled. Importantly (as we will see at the end of this section), the number of meanings that have to be, in principle, expressible is much larger than the number of permitted utterances transmitted between teacher and learner.

When having to produce an utterance, meanings to be conveyed are selected randomly from a fixed, global pool and structured following an “Agent-Patient-Predicate” syntax. For example, a meaning could be structured as  $\langle \text{Agent} = \text{Henk} \rangle$ . When a teacher has no specified utterance associated with a meaning to be expressed, a new utterance has to be invented. This is done by generating arbitrary-length, random character substrings for the components in the meaning that are unknown, while the known components are simply filled in. For example, if a teacher would utter  $\langle \text{Patient} = \text{Ingrid} \rangle$  as “ingrid,”  $\langle \text{Predicate} = \text{Kust} \rangle$  as “kust,”<sup>9</sup> but have no way to express the meaning  $\langle \text{Agent} = \text{Henk} \rangle$ , the final utterance might become something as “fmguhba kust ingrid.” Agents are initially naive. Therefore, when starting the simulation, all utterances produced by the first agent in the chain are completely random.

<sup>9</sup> Dutch third-person singular for “to kiss.”

Learners do not simply internalize what they hear when they get exposed to utterances, but they are able to make generalizations. The precise details of these generalizations can vary from study to study, but what is more important is that they all lead to similar long-term dynamics as meanings get conveyed from agent to agent over long time periods. To summarize, all strategies follow the principle of substituting many complex rules with fewer simpler ones, when there is similarity between the complex rules. More precisely: when multiple meanings are expressed by similar utterances (i.e., they have similar syntax subtrees), a new rule is generated that substitutes the lower-level syntax instance with a more general, higher-level rule. This rule is then applicable to many meaning-to-utterance conversions. These more general, higher-level rules thus enable the agent to generate *compositional* syntax.

Having laid out this framework of cultural evolution of language, the results provided some interesting observations. Over many transmissions, the size of the grammar (i.e., the number of rules converting meaning to utterance) and the number of expressible meanings were recorded. These measurements showed three distinct phases (Fig. 10.2).

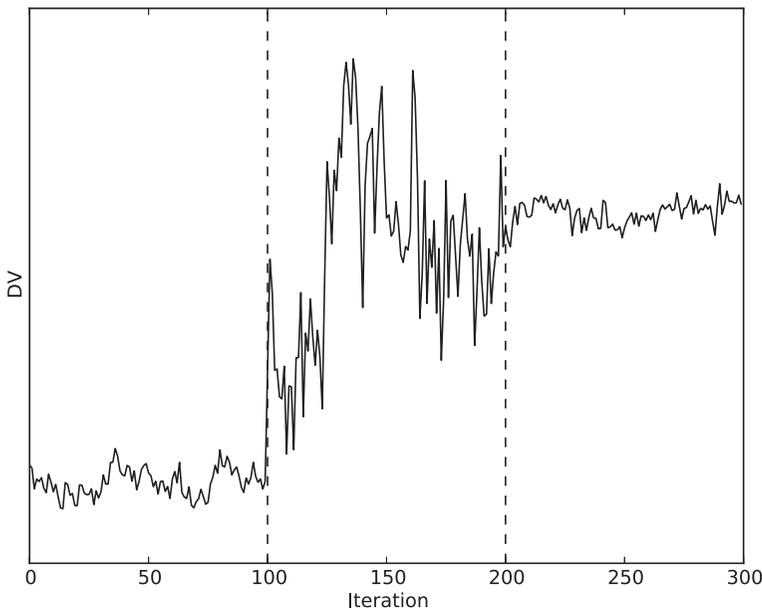


Figure 10.2 A generalization of the phases observable in a typical IL run. Shown is some dependent variable (DV), such as the number of expressible meanings, over time. The second phase (Iteration 100–200) shows a chaotic phase transition following a first phase of stabler dynamics, in turn leading to a state of (semi-)stationary convergence.

In the first phase, both the number of expressible meanings and the size of the grammar remain small, with minor fluctuations over time. Inspection of the agents' syntax trees show that they are completely flat. This is also shown when inspecting the actual utterances used by the agents. These are completely arbitrary and random. For example, the meaning (Agent = Henk, Patient = Ingrid, Predicate = Kust) could be expressed by the utterance "ddadbbbabeedaeeee," while a similar sentence (Agent = Sjaak, Patient = Ingrid, Predicate = Kust) could be expressed by the utterance "d." In other words, utterances are completely idiosyncratic. Each meaning is coupled with a unique utterance and vice versa, and there is no underlying, general structure to them. In many ways, this first phase can be likened to a *proto-language* as hypothesized to be used by earlier hominids using idiosyncratic grunts or gestures to convey meaning (Dediu & Levinson 2013).

The second phase marks a period of large fluctuations in grammar and meaning size, following a brief burst of rapid inflation. It appears there now exist syntactic categories. Some meaning components might now be regularly expressed by the same set of characters, although this is only true for a small number of them. This is reflected in the syntax trees the agents use, which are no longer completely flat, but show occasional branching (the language is said to be partially compositional). Overall, due to the large fluctuations, it is hard to understand what is precisely going on in this phase. One could argue, from an optimization algorithm perspective, that the agents are exploring a search space, briefly occupying some local optima, none of them stable enough to lead to terminal convergence. From a dynamical systems perspective, this second period is typical of what is known as a *phase transition*, analogous to phase transitions in physical systems, e.g., freezing and melting of liquids and solids, respectively.

Eventually however, the chaotic fluctuations in the second phase settle down in a third, static phase where all meanings are expressible using a small number of grammatical rules. In this phase, the language is largely compositional. The rules the agents use are applicable to many linguistic instances and syntax trees reflect this by showing extensive branching of syntactic categories. Few meanings are any longer expressed idiosyncratically. For example, the meaning component (Agent = Henk) could be consistently expressed by the substring "qasd." Furthermore, the usage of particular types of meaning components (i.e., verbs or nouns, predicates or agents/patients) is reflected in the regular positioning in utterances. This is a clear analogue to the use of a consistent word order in actual human languages. Kirby & Hurford (2002) further illustrated the potential of the IL model by showing emergence of *recursive* syntax, but now following a simplified model with agents embedded in a linear chain. Here, utterances are propagated through the chain, with agents first taking the role of learner followed by subsequently assuming the role of teacher

(Fig. 10.1). In this experiment, utterances were interpreted by a perceptron (a simple type of feedforward artificial neural network), while they were produced using statistical inferencing. Interestingly, in an additional experiment (Kirby & Hurford 2002), the ratio of requested meaning was changed such that some meanings were either often or rarely requested, mirroring a Zipfian distribution (Zipf 1949). In this case, the commonly requested meanings remained idiosyncratic, resisting the transmission bottleneck induced generalizations. This parallels the use of irregulars, like the verb “to be” in English.

How might the emergence of compositionality, and even recursive syntax, follow from simple iterated transmission? Consider how agents need to transmit a large number of meanings, with only a limited number of opportunities. This information bottleneck implies that there is no “time” for learners to exhaustively transmit every meaning if they were expressed by idiosyncratic utterances. In such a case, old idiosyncratic utterances are “forgotten,” continuously requiring the invention of new ones. However, if, instead of idiosyncratic utterances, compositional ones could be transmitted, many more meanings would “fit through” the bottleneck. This is because such utterances are based on a more general grammar, applicable to many meaning instances, instead of on an idiosyncratic grammar, which only describes very few meaning instances. Only the most general rules therefore survive the selective pressure the information bottleneck exerts, because only those general rules, applicable to many linguistic instances, are invoked often enough to be guaranteed transmission. Using Darwinian terminology: the more general linguistic replicators have a larger chance of being summoned, thus they are “fitter,” and thus they are more likely to actually replicate when compared to idiosyncratic replicators. This explains how the complex system of human language can be considered as itself evolving to become learnable, instead of the agents, e.g. human beings, evolving to learn the language. This way, it appears cultural evolution alone can be sufficient for features unique to human language, such as compositionality (Hockett 1960), to emerge.

Altogether, however, one might question the claims of the IL experiments that compositionality and recursion in the resulting languages are the effect of the transmission bottleneck alone. First of all, instead of becoming compositional, a language might equally well converge to a state where every meaning is expressed by a single, simple utterance. In this case the language has become maximally transmissible (learnable) but has not become expressive – the language is essentially one big homonym (this issue will be further discussed in Sections 10.2.2 and 10.2.3). Second, the learning algorithm agents use, whether based on inductive logic, neural networks, Bayesian inferencing (Section 10.2.2), or on some other method, is an essential component of the model that co-determines what attractors the language will converge to. For example, in the case of neural networks, it has been shown that only particular

types of networks lead to this convergence, while others cannot overcome an initially naive state and initiate autocatalysis, or are unable to produce faithful reproductions in noisy environments (Smith 2001). This latter criticism then makes a good case that the IL experiments do not explain glossogenesis – the (evolutionary) origin of language in our species from a non-linguistic state. In effect, IL “merely” explains how e.g., the cognitive apparatus, once it provides adequate functionality, is able to shape the convergence of cultural evolution, while it says little on how this apparatus was evolved, whether it was specifically tuned for language learning by domain-specific adaptation, or whether human agents apply more general learning strategies to communicate with each other (explaining the linguistic features we see today as a result of domain-general *exaptations* – the utilization of existing adaptations already in place for novel purposes). These, in many ways even more theoretical, questions will be further addressed in Section 10.3. For now, it can be said that the IL framework mainly models language change on a glossogenetic level, while *ontogenetics* (i.e., developmental factors that shape the learning algorithm) are of a significant influence but not the main focus.

Now that we have established that the learning algorithm, modeling human cognition, has a large influence on the convergence of language via cultural evolution, we might propose a generalization. Instead of the human brain exerting biases, we could imagine, as mentioned in Section 10.1.1 and 10.1.2, that human anatomy and physiology, even at a relatively low, mechanistic level, might exert such biases comparably well. For example, the flexibility of the tympanic membrane or the tone of muscles attached to the ossicles in the inner ear might impose low-level perceptual biases. Alternatively, during speech production, it is not hard to imagine how the shape of the vocal tract, such as hard-palate curvature or lower vocal tract volume, could make the production of particular speech sounds easier or harder, exerting biases in their own terms. As theorized (Section 10.1.2), biases such as these might get saliently expressed even when very small because they can be amplified when iteratively transmitted from speaker to learner. However, bias amplification rests on models with strong assumptions, the *Bayesian IL* models, addressed in Section 10.2.2.

### 10.2.2 *Bayesian Iterated Learning*

The IL models discussed in Section 10.2.1 explain human language as a system emerging from the interactions between agents, strongly implying that commonly seen linguistic features need not be biologically innate or had to evolve by natural selection. However, a series of follow-ups on the IL experiments casts doubts on these assumptions, while making interesting predictions in their own right. The IL models were based on agents using various algorithms such as ones based on neural networks or inductive logic. However, Griffiths and

Kalish (2007) use agents that use Bayes' theorem (Eq. 10.1) to reason about language features.

$$P(h|d) = \frac{P(d|h)P(h)}{\sum_{h \in H} P(h)P(d|h)} \quad (10.1)$$

In the context of IL, Bayes' theorem (Eq. 10.1) describes how agents derive a distribution of language hypotheses (called the *posterior* distribution, or  $P(h|d)$ ) from the observed data (in the form of a *likelihood distribution*,  $P(d|h)$ ) and some sort of bias (a *prior distribution* on language hypotheses,  $P(h)$ ). This bias could, for instance, represent the neural, anatomical, or genetic biases discussed in Section 10.1.2. (The denominator in Eq. 10.1 denotes a normalizing factor.)

To provide a concrete illustration of how this works, consider an IL chain as described in Section 10.2.1, but now with Bayesian agents transmitting utterances to each other (Fig. 10.1). Suppose an agent would have to decide on word order in the utterances they produce, Subject-Object-Verb (SOV) or Subject-Verb-Object (SVO). If the agent's prior distribution would be uniform, i.e., express a 50% preference for either SVO or SOV, the posterior distribution would be completely determined by what word order is used by a teacher agent, and how likely this data is expected to be under different language hypotheses (for example, perceiving an SVO sentence could imply a 95% likelihood that the teacher agent generated it under an SVO hypothesis). Conversely, if this likelihood distribution would be uniform, the prior would dominate. Of course, these simplified situations would not normally occur and the posterior distribution on language hypotheses would always be a product of both likelihood and prior.

Once a posterior distribution has been established, there still remains the issue of what hypothesis an agent will select. To simplify, two options are available: sampling and maximizing. When agents are sampling, language hypotheses are selected with a chance proportional to their probability in the posterior distribution. When maximizing, the hypotheses with the largest posterior is selected. Interestingly, the selection strategy used has a large effect on the resulting languages. When sampling, languages converge on a (Markov chain's) stationary distribution that exactly mirrors the prior. This has the implication that, if human agents are equivalent to Bayesian samplers, the languages we see today strongly reflect biologically innate predispositions. When maximizing, however, convergence is less well understood, but largely seems to reflect only the *ordering* of the hypotheses in the prior distribution. For example, if the prior distribution would express an 80% bias for SVO word order and a 20% bias for SOV, we would expect SVO to be used more often than SOV, but likely not following the 80–20 ratio.

Kirby, Dowman and Griffiths (2007) emphasized the amplification of weak priors by maximizer populations, i.e., a prior with only a slight edge over the other priors might come to dominate over time. This would imply that strong effects (e.g., common speech patterns of phonemes, “universals”) would not require strong biases (e.g., neural modules). Smith and Kirby (2008) then demonstrated that a population of maximizer agents is the evolutionary stable one over a sampler population, i.e., a maximizer population will resist invasion by a sampler minority, but not the other way around. The reason for this is that maximizers are more certain what other agents are doing (i.e., maximizing, selecting the most likely language hypothesis, just like themselves) while with samplers there is a larger chance agents will choose different hypotheses. Since we know that maximizing is the evolutionary stable strategy, and since we know that maximizing can amplify biases, the authors conclude that, altogether, selection should be neutral with regards to prior strength. In other words, cultural evolution leads to *shielding* of bias strength (Kirby et al. 2007).

While intriguing, we must emphasize that these conclusions make a number of assumptions that could be questioned. First, there is the assumption that a bias’s strength is conspicuously apparent to listeners, and that they all are comparably salient (in other words, that they all have an effect of comparable magnitude on perceived utterances). But we know that many high-level language features are invariant to lower-level details (e.g., allophones). For instance, a large bias that induces prohibitive effort on producing the voiced bilabial stop [b], will have no effect on phonemic intelligibility in Spanish if the biased speaker would simply shift to a voiced labiodental fricative [v]. Second, it is assumed biases have homogeneous peripheral costs, i.e., they are all similar in their effects on an organism’s fitness besides the communicative one. But if a particular weak bias is disproportionately costly with respect to, for instance, the ability to swallow or breathe, that bias is probably still selected against. Thus, the notion of selectively neutrality only applies to the bias’s effects on communicative accuracy. Third, IL assumes that not only the utterances but also associated meanings are observable by agents. However, it is less clear how “meanings,” particularly more abstract ones, can be observed by human speakers. Fourth, it is known that the apparent dichotomy between maximizers and samplers Bayesian IL postulates in reality follows a continuum (Kirby et al. 2007); agents can occupy a position that is intermediate between sampling and maximizing. Such agents (that are thus not perfect samplers) over time approximate the behavior of maximizers, causing the eventual language distribution to mirror only the prior ordering. Finally, it is assumed that language transmission can be likened to vertical transmission in a linear, monadic chain (i.e., each learner is being taught by exactly one teacher) or to a population of infinite size (Griffiths & Kalish 2007) (Fig. 10.1). When situated in heterogeneous, polyadic chains (e.g., learners having two or more

teachers), however, language convergence strongly diverges from the monadic chain behavior.

In Smith (2009), it was shown that a polyadic chain of sampling agents converges to the language that has the largest prior, while languages with weaker priors get suppressed. This runs contrary to the monadic results that show samplers converging on a distribution that exactly mirrors the prior, suggesting a type of conformist dynamic. Furthermore, increasing bottleneck size increases transmission fidelity and promotes convergence on the strongest prior (contrary to a monadic chain, where a smaller bottleneck leads to faster prior expression (Kirby et al. 2007)). Similar deviating results have been shown by Ferdinand and Zuidema (2009). First, they showed that homogeneous polyadic maximizers behave almost the same as monadic maximizers do, while polyadic samplers no longer converge to the prior distribution. This is explained by the observation that a learner might receive data as a product of multiple teacher agents that might have entertained different language hypotheses. In other words, the data learners receive is generated from a “virtual” distribution they have no explicit internal representation for, therefore they can no longer be Bayesian rational. The study also investigated heterogeneous polyadic chains where learners had different prior distributions (e.g., teacher agents had different biases). When agents were sampling, the language eventually reflects the average of multiple, homogeneous sampler runs. When maximizing, things get more obfuscated, as language convergence codepends on prior strength and likelihood structure and distinctiveness.

In Dediu (2008, 2009), complete populations of spatially dispersed agents were investigated, thereby including horizontal information flow instead of the purely vertical or oblique transmission seen in the standard IL chains (Fig. 10.1). Dediu (2008) showed that (in populations of non-Bayesian agents) two kinds of biases had different effects on language convergence. With the “initial expectation” bias (e.g., an innate predisposition for some language features), these biases very soon get overruled by linguistic drift. However, the “rate of learning” bias (e.g., an adaptive tendency to acquire particular language features) behaves more akin to Bayesian monadic samplers in the sense that weak biases get amplified through cultural transmission. Second, in Dediu (2009), actual Bayesian agents were then used in the population model, which showed that agents (whether samplers or maximizers) behave as monadic chains of samplers, contradicting the results from the polyadic chains from (Ferdinand & Zuidema 2009; Smith 2009).

These anomalies resulting from relaxing assumptions imply that care must be taken when generalizing findings obtained from Bayesian IL models. The findings in Dediu (2008, 2009), Ferdinand and Zuidema (2009), and Smith (2009) give strong testament to the issue of how to interpret the notions of Bayesian maximizing and sampling agents and to what extent human agents behave as

them. Even in the simple, monadic Bayesian chains introduced by Griffiths and Kalish (2007), it is hard to generalize conclusions on samplers and maximizers to human “agents” since we cannot assume human beings are Bayesian rational in the first place (Ferdinand & Zuidema 2009). Given this insight, it seems that the claim that cultural evolution of language converges on a distribution that mirrors speakers’ innate biases is definitely too strong. This is further highlighted by the results that show that different social topologies (vertical or oblique monadic or polyadic chains, or horizontal populations of agents) can result in different outcomes of the convergence process, and these results sometimes even contradict each other. Such apparent paradoxes, however, can possibly be explained by noticing that the experiments use, for instance, slightly different learning algorithms. As we have seen, a similar sensitivity is seen in the population architecture the IL models follow, and the question which structure is most powerful and most realistic remains unanswered (Mesoudi & Whiten 2008). We note that this sensitivity to initial conditions is typical of complex systems, implying that one cannot explain cultural evolution of language using only a reductionistic, component-based account, but that one, at some point, needs to include the interactions between said “components.” This leads us to raise the question to what extent predictions made by reductionist models such as Bayesian IL are corroborated by experiments on human subjects, which is discussed in Section 10.2.3.

Many of the anomalies of the Bayesian models could potentially be addressed by extending the hypothesis space the agents use in some way or another. For example, Burkett and Griffiths (2010) modified their original model to allow for multiple, heterogeneous teachers by using a hyperprior (a prior distribution over nested prior distributions; Bernardo & Smith, 2009). Using the same approach, Smith, Tamariz, and Kirby (2013) explicitly showed that compositionality in language depends not only on a requirement for learnability or generalizability but also on expressivity (a concern we raised in Section 10.2.1 and that is also observed in IL studies on human subjects; see Section 10.2.3). Using these modifications, it is theoretically possible to model any cognitive process by using nested hypothesis spaces in a hierarchical configuration (Perfors 2012). However, we argue that this more complete specification of the hypothesis space, powerful as it might be, negates one of the most appealing aspects of Bayesian IL, namely its simplicity and its tractability. In these cases, the use of Bayesian agents, in terms of epistemological comprehensibility, reverts to paradigms that are often regarded as opaque, such as artificial neural networks.

To conclude, the Bayesian IL models give us some strong suggestions on how genetic biases might be expressed in languages, but one should keep in mind that they are strongly reductionist in nature, while describing a system that shows all the characteristics of being dynamical. The prediction that genes

might be shielded from natural selection following cultural evolution of language should therefore probably not be taken as the final word on this topic, although it is an intriguing one nonetheless (see Section 10.3 for another example of shielding). Probably the main take-home message from these studies is that genetic biases can indeed affect the outcome of cultural evolution of language, but in very complex ways, and that bias strength is probably not linearly related to its outcome.

### 10.2.3 *Human Subject Trials and Animal Models*

One obvious way of validating the IL models is by testing if human subjects perform in a similar manner. A follow-up on the computational IL studies employed chains of human “agents” that had to generate and transmit utterances (strings of characters) describing meanings (pictograms that represented colored, moving objects, e.g., a blue, spiralling square) (Kirby, Cornish & Smith 2008). Utterance-meaning pairs were then relayed to a second subject tasked with replicating them, in the process transmitting them to a third learner, etc.

Results showed that the “language” the subjects produced converged to an underspecified state (i.e., many homonyms were used), paralleling some of the computational IL experiments mentioned in Sections 10.2.1 and 10.2.2 by Smith et al. (2013). However, when homonyms were filtered in the transmission line, the utterances became much more expressive while remaining learnable. Again, mirroring the computational IL models, utterances became increasingly structured (i.e., compositional) further down the IL chain. Moreover, since, from the subjects’ perspective, both experiments were of an indistinguishable nature, it was concluded that subjects’ intentions, learning strategies, or linguistic background (e.g., their native language) were not a factor in the outcome. However, we speculate that this latter conclusion (on linguistic background) is based on using homogeneous subject pools with little variation. More specifically, if pools were used that grouped subjects on cultural or ethnic background, language convergence is more likely to be different between groups. If this hypothesis indeed was to be established, it could suggest possible innate/acquired biases.

Similar studies on human subjects (Perfors & Navarro 2011; Perfors & Navarro 2014) showed comparable results, but instead of using pictograms, stimuli consisted of simple squares of different sizes and colours. Multiple trials were conducted where the stimulus space followed a smooth gradient (i.e., stimuli each differed from each other to a similar extent) or a more discontinuous one (i.e., some stimuli were very similar, while others were very different). In the gradient condition, utterances converged to an underspecified state, similar to the pictogram stimuli without the expressivity requirement by Kirby et al.

(2008) and what was emphasized in the Bayesian model by Smith et al. (2013). With the discontinuous stimuli, however, the language used at the end of the IL chain reflected the discrete profile of the stimulus-space, even without any explicit requirement for expressivity. In other words, language convergence is influenced not only by innate (e.g., genetic, neural, anatomical) biases but also by the environment (represented by the stimuli distribution) subjects were situated in. A discontinuous environment (i.e., one that provides, one could say, a template for semantic categorizations and an implicit requirement for expressivity) is then able to prevent language convergence to an underspecified state. Furthermore, while the Bayesian sampler studies predict faster convergence with smaller bottlenecks (Section 10.2.2), Perfors and Navarro (2014) show that a larger bottleneck will reflect the environmental biases more strongly. Indeed, a small bottleneck leads to another instance of convergence to an underspecified state (one could say, learners had not enough information to deduce any requirement for expressivity). Overall then, any notion of (genetic) biases can only be considered in a situated context. Not only is the cultural expression of such biases dependent on the interactions between agents, but the (inanimate) environment itself is a factor as well.

As we have seen, the human subject experiments (Kirby et al. 2008; Perfors & Navarro 2011; Perfors & Navarro 2014), as well as their computational analogues discussed in Sections 10.2.1 and 10.2.2, consider cultural evolution of language by generating some form of utterance to express multiple meanings. When this requirement for expressivity is relaxed, the language is probable to converge to a non-compositional, underspecified state, potentially undermining the predictive power of the IL paradigm. A second point of attention on human subject IL studies is that utterances are defined to be discrete, i.e., decomposable in principle (e.g., a string is composed out of discrete characters). This decomposability might be less inherent to continuous speech. Finally, one could argue that the results rely on an existing predisposition for language processing in *adults*, as was already mentioned in Section 10.2.1. In reality, it is first language acquisition in infants that we should consider to be essential in the replication phase in cultural evolution language.

Results by Verhoef and de Boer (2011) and Verhoef, de Boer, and Kirby (2012) argue that these critiques are refutable by proposing a generalization of the traditional IL setup. Instead of having subjects tasked with conveying meanings using discrete utterances, they were required to replicate sounds with a slide whistle. This way, any implicit linguistic assumptions the subjects might impose on the task were attempted to be negated, while also lessening any inherent compositional structure by using a continuous signal. Following this approach, four chains of ten subjects were investigated, each showing a gradual increase in systematic recombination of signals and their utilization in forming compositional utterances. This was also confirmed by showing that the

normalized distance between subjects' received and produced utterances decreased the more often it was transmitted. Moreover, the Shannon entropy (i.e., the "information density" or "signal uncertainty"; Shannon 1948) of produced utterances decreased likewise, confirming the increasingly compositional nature of signals and their usage of shared elements.

The human subject experiments discussed so far seem to largely corroborate the predictions of the computational IL models. However, we argue that there are notable interpretational differences. First of all, in the computational models, agents had perfect memory. When they heard an utterance they were able to, in principle, replicate it flawlessly. Thus, the information bottleneck exerting selective pressure in cultural evolution is usually regarded as a logistic one: An agent can only transmit as many sounds as the bottleneck provides an opportunity for. However, the experiments with human subjects strongly suggest that the bottleneck is a result of limited memory capacity. Subjects are not able to learn all utterances by heart, even if they were exposed to all of them (Verhoef & de Boer 2011; Verhoef et al. 2012), and they are therefore forced to resort to making generalizations. Insofar, this bias in human subject IL seems to be, for a large part, cognitive in origin.

Intriguingly, although animal communication is often judged to be qualitatively different from human language (Hockett 1960), cultural evolution has also been established in nonhuman animals. A study by Fehér, Wang, Saar, Mitra, and Tchernichovski (2009) investigated "language" convergence in zebra finches, *Taeniopygia guttata*, using an experimental design largely similar to the IL studies. When growing up in isolation, the structure of the songs zebra finches produce is markedly different from those produced by zebra finches interacting with each other, demonstrating unusually long syllable duration, stuttering, and more broadband noise. In a first experiment, four zebra finches grown up in such a socially isolated situation served as tutors for juveniles. The juveniles, however, did not copy the tutors' song structure with high fidelity. Instead, the songs appeared more similar to the wild-type songs. When transmitted in succession from bird to bird, like in the IL experiments, song structure approached the wild type asymptotically. A similar convergence was found when isolates were founding, genetically as well as culturally, small colonies of zebra finches that allowed for horizontal transmission.

Conceptually, the zebra finch study directly addresses the influence of (genetic) biasing in cultural evolution. More specifically, there seems to be an intrinsic bias that forces the zebra finches to converge on the wild-type song structures, even when founded by isolates. This has a number of possible explanations. First, we suggest that the song structures in isolates and the tutored birds is resultant of one and the same bias and that the expression of this bias accumulates the more often it is transmitted (as would be expected from the amplification of weak biases following Bayesian IL using maximizers;

Section 10.2.2). In that case, however, the precise nature of the bias (e.g., anatomical, perceptual) is hard to pin down on face value. Alternatively, however, one might propose that the bias does not manifest in the isolates at all, implying some transmission factor between individuals is responsible for the wild-type convergence. For instance, it might be possible that the zebra finch mimicking behavior introduces a bias that is unused and therefore not expressed when birds develop in isolation. Perceptual biases, again likely not of importance in isolates' song structure, might introduce convergence to wild-type songs equally well. However, production biases (e.g., anatomical) seem less likely, since song production is a factor that is of comparable utilization in both isolates' as tutored birds' songs. Finally, there might be some sort of sexual selection in the colony model, e.g., females are more likely to mate with males that produce wild-type-like songs (although this seems unlikely given the observation that both (single-sex) chain as well as (mixed-sex) colony experiments produce similar (i.e., wild-type songs).

#### 10.2.4 *Self-Organization of Vowel Space*

The IL model (Sections 10.2.1 and 10.2.2) is a powerful approach and its predictions have been supported by human subject as well as animal models (Section 10.2.3). Furthermore, Bayesian IL (Section 10.2.2) allows the precise specification of priors that can be interpreted as representing innate biases with an (ultimately, at least partial) genetic basis. A problem however, as already addressed in Section 10.2.2, is its contradictory, strongly reductionist nature while describing language as an emergent property in a dynamical system. Moreover, the provided precision in defining priors constitutes a double-edged sword. More specifically, when modelling a natural system, the bias is often not known in detail and therefore hard to formalize in a Bayesian way. An alternative strategy is to look at cultural evolution of language from a *self-organizing*, dynamical systems perspective, in many ways resembling exemplar-based phoneme perception models (e.g., Kruschke 1992; Johnson 2005).

A study by de Boer (2000a, 2000b) showed that a self-organizing, population-level system can explain vowel dispersion in human languages accurately. Pairs of agents are iteratively selected to play an "imitation game." One agent initiates the game by transmitting a vowel that the imitator tries to match to one of its internally stored vowel prototypes using the Euclidean distance between formants. The imitator in turn produces a vowel based on the selected prototype, which the initiator in turn matches to a prototype as well. If both agents classify the same prototype, the communication game is a success and the imitator's prototype is shifted toward the perceived vowel. New prototypes are produced following unsuccessful games when the closest match

has a matching history too successful to be discarded. The prototype space is periodically cleansed based on the success history of prototypes.

The results show that, after 200 games, clusters in the vowel space start to form, explained by the fact that agents try to imitate each other while there is (cultural) selective pressure to have a maximum number of maximally distinguishable vowels. Thus, after 2,000 iterations, the vowel space is occupied by a few, tight clusters that are (between-cluster) maximally dispersed. An experiment where agents were periodically removed and new agents were inserted (where younger agents were able to change their vowel repertoire more easily, modeling first language acquisition in infants) showed similar results. Promisingly, the vowel clusters to which the experiments converged showed striking similarities with those seen in human languages (cf. Schwartz, Boë, Vallée & Abry 1997). Later studies used a similar self-organization approach, now including a temporal dimension (Zuidema & de Boer 2009; de Boer & Zuidema 2010). In doing so, it was shown that common sounds were reused but at different points in time. In effect, this strategy maximizes dispersion by using the temporal dimension as a distinguishability axis, leading to the development of (a system analogous to) combinatorial phonology that is often cited as unique to human language (Hockett 1960). Thus again, it appears some language properties can be explained without any need for nativism or biological evolution.

Although designed as a self-organizing system, the model by de Boer (2000a, 2000b) still incorporates an explicit, procedural definition of the language games agents play, both in terms of how agents communicate with each other as well as how they organize their phonemes internally. Moreover, the study therefore relies on the assumption that agents had the a priori capacity to communicate and that they experienced an explicit pressure to exploit this capacity. Thus, similarly to the IL models described earlier (Section 10.2.1), the studies do not explain glossogenesis, but are confined to the domain of cultural evolution of language when the capacity for it is already established. To address this, Oudeyer (2005a, 2005b) used an approach similar to de Boer (2000a, 2000b) but now basing agents' internals on self-organizing maps (SOMs) (Kohonen 1982, 2001). SOMs are often used for the purpose of dimensional reduction (akin to multidimensional scaling) and they are therefore applicable to map a high-dimensional input vector (e.g., the first few formant frequencies of a waveform) to some internal representation of fewer dimensions, argued to be biologically realistic. Subsequently, they can also be used to map an internal, high-dimensional representation onto a lower-dimensional output vector. More technically, an SOM consists of a layer of parallel neurons that each express a sensitivity for a particular input vector, where activation in one neuron will "bleed over" to neighboring ones (e.g., following a Gaussian distribution). Critically, these maps, as the name suggests, self-organize in that those neurons that yield the highest activation to an input vector tune themselves to respond to that

vector even more strongly. Following this architecture, Oudeyer (2005b) uses two of these SOMs, one perceptual and one motor map, in series to transform speech sounds into articulatory gestures. For this purpose, the two maps are fully connected and these connections are updated using a Hebbian learning rule (Hebb 1949). Pairs of agents are randomly selected: one produces a sound that the other one hears. Furthermore, the study used quasi-realistic<sup>10</sup> vowel production and perception models, similar to those used in de Boer (2000a, 2000b).

In a first experiment (Oudeyer 2005a), however, vowel production was simplified, i.e., the mapping from the articulatory gestures to acoustics was linear. This nevertheless demonstrated that agents interacting with each other self-organize their vowel space around a few attractors, what could be called analogues to the prototypes in de Boer (2000a, 2000b). Interestingly however, this organization also emerged when agents were only allowed to self-talk, as in the babbling of infants. Expectedly, agents did not converge on a shared vowel dispersion pattern in this scenario, as was the case when agents were interacting with each other.

The experiment, while showing clear vowel convergence, did not reflect a vowel distribution seen in actual human populations. The second experiment (Oudeyer 2005a), however, using the more realistic gesture-to-vowel mapping, illustrated a more convincing picture in that respect: The distribution of vowel frequencies (i.e., how many vowel “phonemes” emerged in the simulations) clearly approximated that seen in human populations (Ladefoged & Maddieson 1998). Moreover, the dispersion in the vowel space showed a close resemblance to human language as well. Furthermore, in the follow-up (Oudeyer 2005b), combinatoriality (i.e., the reuse of segments at different points in time) in the vowel system was demonstrated by extending the model with a temporal neural map that self-organized through pruning the less active neurons (simulating, in a way, apoptosis – programmed cell death), allowing the production of (small and indeed combinatorial) sequences of vowels.

While the convergence on a discrete, one could claim phonological, vowel system occurs without the explicit intention to communicate, and even without any contact with peers whatsoever, the degree of realism of the “vocal tract” (linear versus parameterized) leads to different outcomes of the self-organizing process, given a tentative example of anatomical biasing. This issue was further

<sup>10</sup> We use the term “quasi,” since the suggested realism is derived from numerical transformations. More precisely, vowels are produced using three parameters: lip rounding, tongue height, and tongue fronting. Vowel perception is based on a modified Barks transformation (Zwicker 1961) that accounts for relative narrow-band, high-frequency perceptual indiscriminability in human subjects. Thus, there is no simulated “physicality” in these models, but they remain relatively abstract, quasi-realistic instead.

emphasized in Oudeyer (2005a) that introduced a (“metabolic”) energy cost on vocal tract “displacements,” again resulting in different vowel dispersion patterns. None of these dispersion patterns were realistic, however. This, the author concludes, could have been expected, since dynamical systems are notoriously sensitive to initial conditions and perturbations of any kind. Failing to accurately model even a single biasing factor might therefore result in totally different outcomes. This realization then raises questions on the self-organization approach, particularly on what level of abstraction would be appropriate when modeling human language, i.e., with the aim to make testable predictions. More concretely, how would one estimate whether all relevant (genetic, anatomical, or other) biasing factors are accounted for while keeping the model, following Occam’s razor, as simple as possible? These questions remain as of yet unanswered.

Overall, and more theoretically, the self-organizing models describe a form of ontogenetic development – *phenotypic plasticity* – that can be conceptualized to work in conjunction with natural selection. It is not hard to imagine how the search space of human language features and parameters is enormous, and that natural selection in isolation is possibly not powerful enough to explore this space efficiently (Ball 1999). In that regard, ontogenetics such as self-organizing processes might actually amplify natural selection. For example, the studies by de Boer (2000a, 2000b) showed that language-like self-organization processes took place for a large range of parameters that determined the behavior of the SOMs. Thus, the target volume natural selection has to find to let language emerge is drastically inflated by application of this “local search.” In more ethological terms, this can be visualized by imaging natural selection traversing a search space of phenotypes in a relatively slow but robust fashion (i.e., emphasizing specialism), while ontogenetics allows for lifetime adaptations (i.e., emphasizing generalism) (Turney 1996). The realization that phenotypic plasticity might play a major role in biological evolution will probably have a massive effect on our understanding of evolutionary theory (Pigliucci 2007). For instance, this interaction might lead to another example of a situation where genes are being shielded from natural selection (since ontogenetics allow for a wider range of genetic polymorphisms to be effective). These interactions between phylogenetics, glossogenetics, and ontogenetics is discussed further in Section 10.3.

### 10.3 Cultural Feedback

So far, we have mainly discussed how genetics might bias cultural evolution of language, acting through neurocognition and the anatomy and physiology of the production and perception systems. This is a simplification of the natural

situation, however. For example, not only does human physiology shape language on a cultural level, but certain language features might conceivably become *assimilated* into the genome, and this genome, in turn, shapes language acquisition, processing, production, and perception. This thus effectively describes a dynamical system that would consist of three layers coupled in a feedback loop: genes, physiology, and culture.

The Baldwin Effect (Baldwin 1896) has generated considerable interest, describing the interaction between genes and phenotypic plasticity. First, the effect postulates that organisms might evolve to a state predisposed to adaptability. For example, organisms might develop complex nervous systems that allow them to cope with a dynamic environment. Secondly, the effect proposes that such ontogenetically acquired traits can be internalized into the organism's genome. Thus, even if we observe language modules or other biasing candidates in human anatomy and physiology, we cannot conclude a causal role. On the contrary, it might very well be the case that culturally expressed language features have caused the development of these (physiological) traits, instead of the other way around.

Investigations on the Baldwin Effect have a history of computational modelling. Hinton and Nowlan (1987), for instance, show that the use of phenotypic plasticity enables an organism to explore a search space of phenotypes and how, as mentioned in Section 10.2.3, this can be considered a form of local search on top of natural selection. Once optima are ontogenetically found, they increase an organism's fitness, after which they can subsequently be internalized into the genome, trading flexibility (generality) for optimality (specificity) when the situation requires. However, more recent experiments have cast doubts on the assimilation of language features expressed on a cultural level. For example, there are strong suggestions that in certain situations, described in Sections 10.2.3 and 10.2.4, ontogenetics or particular convergence patterns following cultural evolution might lead genes to be shielded from natural selection/assimilation.

In the domain of language evolution, the Baldwin Effect is often regarded from the *niche construction* interpretation, which states that the expression of stable language features (a stable cultural niche) exerts selective pressure on individuals to assimilate those features (Deacon 1997; Odling-Smee, Laland & Feldman 2003). Eventually, the predispositions to express particular language features might become so strong they appear "innate" (i.e., they are developed before birth, but the concept of innateness is notoriously complex; Mameli & Bateson 2006), via a process known as canalization (Waddington 1942). Such mechanisms might explain many cases of gene-culture coevolution such as the persistence of the adaptation to digest fresh milk in populations with dairy traditions (Richerson & Boyd 2008; Laland, Odling-Smee & Myles 2010; Richerson, Boyd & Henrich 2010; Richerson & Christiansen 2013).

The Baldwin Effect is conceptually closely related to the evolutionary mechanisms of adaptation and exaptation (Gould & Vrba 1982). To reiterate, adaptation describes the evolution of a novel, domain-specific trait (e.g., teeth for mastication), while exaptation relates to the co-option of existing traits for purposes they were not originally adapted for. An example of exaptation is the co-opting of feathers that some theropod dinosaurs used for thermo-regulation but with slight modifications were utilized to aid in flight in, e.g., birds (Ostrom 1976). Another example of exaptation might be the use of the lungs to produce vocalizations in animal communication or, more speculatively, of domain-general pattern recognition capabilities of the human brain in language cognition (Christiansen & Chater 2008). This latter example is particularly relevant because it is largely in line with the IL models discussed in Section 10.2, which describe how agents endowed with only general learning algorithms are able to produce complex languages without requiring any language modules.

As mentioned earlier in this section, some degree of modularity of mind might still arise out of selective pressure from sufficiently stable linguistic niches, leading to a possible genetic encoding to express the corresponding linguistic features. The required stability of this assimilation was investigated by Baronchelli, Chater, Christiansen, and Pastor-Satorras (2013), who used a particle model to run simulations on populations of “generalist” and “specialist” individuals. They showed that specialist individuals only evolved when they were confronted with a situation that allowed for little genetic and environmental change. When environmental change was larger, generalist individuals were favored. Drawing a parallel to language-gene coevolution, because language changes fast, this might imply that generalist speakers are favored, i.e., those that do not evolve language-specific adaptations.

Preluding the more general, abstract findings by Baronchelli et al. (2013), Chater, Reali, and Christiansen (2009) demonstrated that the assimilation of language features indeed requires a low rate of linguistic change in a computer simulation where language features exert selective pressure on the genome. More specifically, in the (unrealistic) scenario where the rate of linguistic change was equal to the rate of genetic change, assimilation is already substantially reduced when compared to when language was completely stable (let alone if the rate of linguistic change is higher). When the language was in turn partially (i.e., 50%) genetically determined, this provided a stabilizing influence that increased the assimilation rate. However, this ratio of genetic determinism appeared to be so (again, unrealistically) high that no distinction could be made between having a selective pressure from language and having no selective pressure at all (i.e., random genetic drift). In other words, the 50% genetic determinism scenario describes a situation in which the language can, for all intents and purposes, be said to be complete genetically determined, a situation that is extremely unlikely. Furthermore, high linguistic change favoured

the evolution of “neutral” alleles that allowed for more flexibility in expressing language features, while slow linguistic change lead to the evolution of alleles expressing such language features by genetic determinism (Baronchelli, Chater, Pastor-Satorras & Christiansen 2012). A similar finding was found when multiple populations were simulated that had interlingual contact: When an individual’s fitness was codetermined by its ability to learn a foreign population’s language, this again lead to the evolution of neutral alleles. Finally, the study also showed that features assimilated during a phase of slow linguistic change (e.g., during a proto-language) mutate into neutral alleles when the rate of change increases. Thus, the authors conclude that it is unlikely that assimilated remnants of a proto-language we might have spoken in the past still reside in our genome.

While interesting, most models on the Baldwin Effect make a number of simplifying assumptions. First of all, it is assumed that genes isomorphically correspond with linguistic features and that all these features had equal weight in expressing meaning. However, given what we know about the genetic bases of language (Fisher 2006), this assumption is almost certainly false. A second simplification is that all cultural linguistic features are equally stable. However, this has been long demonstrated not to be the case for human language (e.g., Maddieson 1984; Schwartz et al. 1997; Dunn et al. 2011), having lead to the suggestion of “universals” in the first place. Stable features such as these are of course more likely candidates for assimilation than unstable ones or those the stability of which is homogeneous and indistinct (such as in Chater et al. 2009, where this was recognized but not further addressed). Finally, the studies assume a simple, linear quantification of the cost of flexibility, i.e., the more plastic an organism, the longer it will take to arrive at the right phenotype. However, this quantifies only one of a number of cost associations with learning (Mayley 1996); none of them likely to have a linear signature.

To summarize, although the validity of Baldwinian niche construction with respect to cultural evolution of language is still debated, with multiple authors contradicting each other, we should consider that apparent domain-specific language modules or other adaptations, even if they were conclusively demonstrated, still do not imply the biological evolution of language-specific adaptations. Following this reasoning, if one were to discover apparent biasing factors, it might be the case that these only evolved after some stable linguistic feature became expressed, i.e., conforming to the selective pressure exerted from language itself instead of the other way around.

## 10.4 Conclusion

Biological and cultural evolution show striking similarities (Section 10.1.2). While not exact homologues, the principles of replication, variation, and

selection are found in both domains. This has the potential to explain linguistic features without the need to invoke “universals,” “language modules,” or the (purely) biological evolution of language. However, we are by no means stating that human languages do not, to some extent, show a degree of semi-universality, that certain cortical areas are more involved in language processing than others, or that biological natural selection is of no importance to explain the complexities of language. Nevertheless, the concept of cultural evolution is a large explanatory factor in itself and should be considered to interact in close conjunction with biological and physiological mechanisms. This insight is a shift away from the dominant, cognitivist perspective of speech perception and production, instead considering human language to (weakly) emerge from the interactions between speakers. Importantly, none of these speakers have any a priori, explicit internalization or predisposition to express linguistic features, but are exerting a social force on language convergence, shaping it in subtle, nonspecific ways. Extra-linguistic factors as well as innate biases speakers might have are therefore of importance, but we cannot hope to fully understand this convergence when taking a full reductionist approach is isolation.

Two frameworks are often used in computational modeling of cultural evolution. Iterated learning provides the benefit that it makes accurate, strong predictions. The “standard” IL experiments have, for instance, demonstrated the emergence of compositionality, recursion and the appearance of irregulars in Zipfian distributions (Section 10.2.1). Furthermore, using Bayesian agents in an IL framework provides a precise, tractable specification of biasing factors and, in doing so, illustrates how certain learning strategies might result in an amplification of weak biases, eventually hypothesized to lead to selective neutrality (or “shielding”) of bias strength (Section 10.2.2). Furthermore, the results are supported by trials on human subjects and by animal models (Section 10.2.3). However, as all models, IL forms an abstraction from the real world and rests on numerous assumptions. Relaxing those expectedly weakens some of the aforementioned predictions.

An alternative to using IL is to take the complex systems route, emphasizing self-organization to an even greater extent than IL and claiming that, for example, vowel space dispersion can be explained from the interactions between agents that might not even have any a priori conception of language, emphasizing glossogenesis that IL does not address (Section 10.2.4). More conceptually, these models illustrate how developmental, self-organizing processes can amplify the ability for natural selection to explore the vast search space of phenotypes, adding a degree of flexibility that selection alone does not provide. However, they also explicitly show that, an observation also relevant for IL, these systems are very sensitive to initial conditions and on-the-fly perturbations. In the end, our aim is to obtain a model that is as simple as possible,

without losing predictive power. This observed sensitivity of complex systems then questions what the appropriate level of abstraction for modeling cultural evolution of language is.

Extending this question from the topic of genetic biasing to gene-language coevolution raises more, similar issues (Section 10.3). It has long been recognized that ontogenetically acquired traits can be internalized into the genome. Language features have often been proposed to be candidates for this assimilation following Baldwinian niche construction, where plastic flexibility is traded for genetic rigidity when the environment is sufficiently stable. Computer models, however, question the feasibility of this hypothesis, showing that language features have to be very stable indeed or they need to be, for an unreasonable part, genetically determined for assimilation to occur. Once more, however, these models make a number of nontrivial assumptions, casting doubt on the validity of their level of abstraction. Overall, they, together with the divided literature from the biological sciences, illustrate that assimilation cannot be ruled out and therefore that, even if we would establish something like a cortical language module encoding word order, it would not imply its biological evolution as a causal factor. Instead, it might be that a (stable) cultural feature assumes that role and that any apparent adaptations are, upon close inspection, mere exaptations of existing traits.

To conclude this chapter, we hope to have shown the potential for using computer modeling to investigate the evolution of language as multiple, interacting domains subject to similar Darwinian principles, emphasizing the role of cultural evolution and the biasing effects genetics might have. While many questions remain to be answered, we regard this approach to be essential to fully account for the vast richness of human language and encourage a further transcendence of traditional disciplinary boundaries in this endeavour.

## Acknowledgments

This chapter was funded by VIDI grant 276-70-022 of the Netherlands Organisation for Scientific Research (NWO) as part of the “Genetic Biases in Languages and Speech” research project. The authors are supported by the Max Planck Society.

## References

- Baldwin, J. M. (1896), ‘A new factor in evolution’, *American naturalist* pp. 536–553.
- Ball, P. (1999), *The self-made tapestry: Pattern formation in nature*, Oxford University Press.
- Baronchelli, A., Chater, N., Christiansen, M. H. & Pastor-Satorras, R. (2013), ‘Evolution in a changing environment’, *PLoS one* **8**(1), e52742.

- Baronchelli, A., Chater, N., Pastor-Satorras, R. & Christiansen, M. H. (2012), 'The biological origin of linguistic diversity', *PLoS one* **7**(10), e48029.
- Bernardo, J. M. & Smith, A. F. (2009), *Bayesian theory*, Vol. 405, John Wiley & Sons.
- Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., Gray, R. D., Suchard, M. A. & Atkinson, Q. D. (2012), 'Mapping the origins and expansion of the Indo-European language family', *Science* **337**(6097), 957–960.
- Burkett, D. and Griffiths, T. L. (2010), 'Iterated learning of multiple languages from multiple teachers', *The evolution of language: Proceedings of Evolang* pp. 58–65.
- Butcher, A. (2006), *Australian Aboriginal languages: Consonant-salient phonologies and the 'place-of-articulation imperative'*, New York and Hove: Psychology Press, pp. 187–210.
- Calvin, W. H. (2002), *A brain for all seasons: Human evolution and abrupt climate change*, University of Chicago Press.
- Campbell, L. & Poser, W. J. (2008), *Language classification: History and method*, Cambridge University Press.
- Carroll, S. B. (2005), *Endless forms most beautiful: The new science of evo devo and the making of the animal kingdom*, number 54, W.W. Norton & Company.
- Chater, N., Reali, F. & Christiansen, M. H. (2009), 'Restrictions on biological adaptation in language evolution', *Proceedings of the National Academy of Sciences* **106**(4), 1015–1020.
- Chomsky, N. (1965), *Aspects of the theory of syntax*, number 11, MIT press.
- Chomsky, N. (1986), *Knowledge of language: Its nature, origin, and use*, Greenwood Publishing Group.
- Christiansen, M. H. & Chater, N. (2008), 'Language as shaped by the brain', *Behavioral and Brain Sciences* **31**(05), 489–509.
- Croft, W. (2000), *Explaining language change: An evolutionary approach*, Pearson Education.
- Dávid-Barrett, T. & Dunbar, R. (2013), 'Processing power limits social group size: Computational evidence for the cognitive costs of sociality', *Proceedings of the Royal Society B: Biological Sciences* **280**(1765).
- Dawkins, R. (1976), *The selfish gene*, Oxford University Press.
- de Boer, B. (2000a), 'Emergence of vowel systems through self-organisation', *AI Communications* **13**(1), 27–39.
- de Boer, B. (2000b), 'Self-organization in vowel systems', *Journal of Phonetics* **28**(4), 441–465.
- de Boer, B. & Fitch, W. T. (2010), 'Computer models of vocal tract evolution: An overview and critique', *Adaptive Behavior* **18**(1), 36–47.
- de Boer, B. & Zuidema, W. (2010), 'Multi-agent simulations of the evolution of combinatorial phonology', *Adaptive Behavior* **18**(2), 141–154.
- Deacon, T. (1997), *The symbolic species: The co-evolution of language and the brain*, number 202, WW Norton & Company.
- Dediu, D. (2008), 'The role of genetic biases in shaping the correlations between languages and genes', *Journal of Theoretical Biology* **254**(2), 400–407.
- Dediu, D. (2009), 'Genetic biasing through cultural transmission: Do simple Bayesian models of language evolution generalise?', *Journal of Theoretical Biology* **259**(3), 552–561.

- Dediu, D. (2011), 'Are languages really independent from genes? If not, what would a genetic bias affecting language diversity look like?', *Human Biology* **83**(2), 279–296.
- Dediu, D. & Levinson, S. C. (2013), 'On the antiquity of language: The reinterpretation of Neandertal linguistic capacities and its consequences', *Frontiers in Psychology* **4**.
- Dunn, M., Greenhill, S. J., Levinson, S. C. & Gray, R. D. (2011), 'Evolved structure of language shows lineage-specific trends in word-order universals', *Nature* **473**(7345), 79–82.
- Everett, C., Blasi, D. E. & Roberts, S. G. (2015), 'Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots', *Proceedings of the National Academy of Sciences* **112**(5), 1322–1327.
- Farrell, S., Wagenmakers, E.-J. & Ratcliff, R. (2006), '1/f noise in human cognition: Is it ubiquitous, and what does it mean?', *Psychonomic Bulletin & Review* **13**(4), 737–741.
- Fehér, O., Wang, H., Saar, S., Mitra, P. P. and Tchernichovski, O. (2009), 'De novo establishment of wild-type song culture in the zebra finch', *Nature* **459**(7246), 564–568.
- Ferdinand, V. and Zuidema, W. (2009), Thomas' theorem meets Bayes' rule: A model of the iterated learning of language, in 'Proceedings of the 31st Annual Conference of the Cognitive Science Society', Cognitive Science Society Austin, TX, pp. 1786–1791.
- Fisher, S. E. (2006), 'Tangled webs: Tracing the connections between genes and cognition', *Cognition* **101**(2), 270–297.
- Fodor, J. A. (1983), *The modularity of mind: An essay on faculty psychology*, MIT Press.
- Gould, S. J. & Vrba, E. S. (1982), 'Exaptation – a missing term in the science of form', *Paleobiology*, pp. 4–15.
- Gray, R. D. & Atkinson, Q. D. (2003), 'Language-tree divergence times support the Anatolian theory of Indo-European origin', *Nature* **426**(6965), 435–439.
- Griffiths, T. L. and Kalish, M. L. (2007), 'Language evolution by iterated learning with bayesian agents', *Cognitive Science* **31**(3), 441–480.
- Hamilton, W. D. (1963), 'The evolution of altruistic behavior', *American Naturalist* pp. 354–356.
- Hanke, D. (2004), 'Teleology: The explanation that bedevils biology', *Explanations: Styles of Explanation in Science*, pp. 143–155.
- Hebb, D. O. (1949), *The organization of behavior: A neuropsychological approach*, John Wiley & Sons.
- Henrich, J. & McElreath, R. (2003), 'The evolution of cultural evolution', *Evolutionary Anthropology: Issues, News, and Reviews* **12**(3), 123–135.
- Hinton, G. and Nowlan, S. (1987), 'How learning can guide evolution', *Complex Systems* **1**(1), 495–502.
- Hockett, C. (1960), 'The origin of speech', *Scientific American* **203**, 88–96.
- Johnson, K. (2005), Speaker normalization in speech perception, in *The handbook of speech perception*, John Wiley & Sons, pp. 363–389.
- Kirby, S., Cornish, H. & Smith, K. (2008), 'Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language', *Proceedings of the National Academy of Sciences* **105**(31), 10681–10686.

- Kirby, S., Dowman, M. and Griffiths, T. L. (2007), 'Innateness and culture in the evolution of language', *Proceedings of the National Academy of Sciences* **104**(12), 5241–5245.
- Kirby, S. & Hurford, J. R. (2002), The emergence of linguistic structure: An overview of the iterated learning model, in *Simulating the evolution of language*, Springer, pp. 121–147.
- Kohonen, T. (1982), 'Self-organized formation of topologically correct feature maps', *Biological Cybernetics* **43**(1), 59–69.
- Kohonen, T. (2001), *Self-organizing maps*, Vol. 30, Springer.
- Kröger, R. H. & Biehler, O. (2009), 'Space-saving advantage of an inverted retina', *Vision Research* **49**(18), 2318–2321.
- Kruschke, J. K. (1992), 'Alcove: An exemplar-based connectionist model of category learning', *Psychological Review* **99**(1), 22.
- Ladd, D. R., Dediu, D. & Kinsella, A. R. (2008), 'Languages and genes: Reflections on biolinguistics and the nature-nurture question', *Biolinguistics* **2**(1), 114–126.
- Ladefoged, P. (1984), Out of chaos comes order? Physical, biological, and structural patterns in phonetics, in A. Cohen & M. van den Broecke, eds, *Proceedings of the Tenth International Congress of Phonetic Sciences*, Foris Publications: Dordrecht, Holland, pp. 83–95.
- Ladefoged, P. & Maddieson, I. (1998), 'The sounds of the world's languages', *Language* **74**(2), 374–376.
- Laland, K. N., Odling-Smee, J. & Myles, S. (2010), 'How culture shaped the human genome: Bringing genetics and the human sciences together', *Nature Reviews Genetics* **11**(2), 137–148.
- Levinson, S. C. & Gray, R. D. (2012), 'Tools from evolutionary biology shed new light on the diversification of languages', *Trends in Cognitive Sciences* **16**(3), 167–173.
- Lin, C. & Shu, F. H. (1964), 'On the spiral structure of disk galaxies', *The Astrophysical Journal* **140**, 646.
- Maddieson, I. (1984), *Patterns of sounds*, Cambridge University Press.
- Mameli, M. & Bateson, P. (2006), 'Innateness and the sciences', *Biology and Philosophy* **21**(2), 155–188.
- Mayley, G. (1996), The evolutionary cost of learning, in *Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, pp. 458–467.
- Mesoudi, A. & Whiten, A. (2008), 'The multiple roles of cultural transmission experiments in understanding human cultural evolution', *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**(1509), 3489–3501.
- Miller, G. (2001), 'The mating mind: How sexual choice shaped the evolution of human nature', *Psychology* **12**(8), 1–15.
- Odling-Smee, F. J., Laland, K. N. & Feldman, M. W. (2003), *Niche construction: The neglected process in evolution*, number 37, Princeton University Press.
- Okasha, S. (2006), *Evolution and the Levels of Selection*, Vol. 16, Clarendon Press Oxford.
- Ostrom, J. H. (1976), 'Archaeopteryx and the origin of birds', *Biological Journal of the Linnean Society* **8**(2), 91–182.
- Oudeyer, P.-Y. (2005a), 'The self-organization of combinatoriality and phonotactics in vocalization systems', *Connection Science* **17**(3–4), 325–341.

- Oudeyer, P.-Y. (2005b), 'The self-organization of speech sounds', *Journal of Theoretical Biology* **233**(3), 435–449.
- Pagel, M., Atkinson, Q. D. & Meade, A. (2007), 'Frequency of word-use predicts rates of lexical evolution throughout Indo-European history', *Nature* **449**(7163), 717–720.
- Perfors, A. (2012), 'Bayesian models of cognition: What's built in after all?', *Philosophy Compass* **7**(2), 127–138.
- Perfors, A. & Navarro, D. J. (2011), Language evolution is shaped by the structure of the world, in *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, Cognitive Science Society.
- Perfors, A. and Navarro, D. J. (2014), 'Language evolution can be shaped by the structure of the world', *Cognitive Science* **38**(4), 775–793.
- Pigliucci, M. (2007), 'Do we need an extended evolutionary synthesis?', *Evolution* **61**(12), 2743–2749.
- Pinker, S. & Bloom, P. (1990), 'Natural language and natural selection', *Behavioral and Brain Sciences* **13**(4), 707–727.
- Richerson, P. J. & Boyd, R. (2008), *Not by genes alone: How culture transformed human evolution*, University of Chicago Press.
- Richerson, P. J., Boyd, R. & Henrich, J. (2010), 'Gene-culture coevolution in the age of genomics', *Proceedings of the National Academy of Sciences* **107**(Supplement 2), 8985–8992.
- Richerson, P. J. & Christiansen, M. H. (2013), *Cultural evolution: Society, technology, language, and religion*, MIT Press.
- Schwartz, J.-L., Boë, L.-J., Vallée, N. & Abry, C. (1997), 'Major trends in vowel system inventories', *Journal of Phonetics* **25**(3), 233–253.
- Shannon, C. E. (1948), *The mathematical theory of communication*, University of Illinois Press.
- Smith, K. (2001), The evolution of learning mechanisms supporting symbolic communication, in *CogSci2001, the 23rd Annual Conference of the Cognitive Science Society*, Citeseer.
- Smith, K. (2009), Iterated learning in populations of bayesian agents, in *Proceedings of the 31st annual conference of the cognitive science society*, Austin, TX: Cognitive Science Society, pp. 697–702.
- Smith, K. and Kirby, S. (2008), 'Cultural evolution: Implications for understanding the human language faculty and its evolution', *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**(1509), 3591–3603.
- Smith, K., Tamariz, M. & Kirby, S. (2013), Linguistic structure is an evolutionary trade-off between simplicity and expressivity, in *Proceedings of Cogsci 2013*, pp. 1348–1353.
- Turney, P. (1996), Myths and legends of the Baldwin Effect, in *Proceedings of the Workshop on Evolutionary Computing and Machine Learning at the 13th International Conference on Machine Learning*, pp. 135–142.
- Verhoef, T. & de Boer, B. (2011), Cultural emergence of feature economy in an artificial whistled language, in *Proceedings of the 17th international congress of phonetic sciences. Hong Kong: City University of Hong Kong*, pp. 2066–2069.
- Verhoef, T., de Boer, B. & Kirby, S. (2012), Holistic or synthetic protolanguage: Evidence from iterated learning of whistled signals, in *The evolution of language:*

- Proceedings of the 9th international conference (EVOLANG9)*, World Scientific, pp. 368–375.
- Waddington, C. H. (1942), ‘Canalization of development and the inheritance of acquired characters’, *Nature* **150**(3811), 563–565.
- Williams, G. C. (1966), *Adaptation and natural selection: A critique of some current evolutionary thought*, Princeton University Press.
- Wilson, D. S. & Wilson, E. O. (2008), ‘Evolution “for the good of the group”’, *American Scientist* **96**(5), 380–389.
- Wynne-Edwards, V. C. (1962), *Animal dispersion in relation to social behaviour*, Hafner Pub. Co.
- Wynne-Edwards, V. C. (1986), *Evolution through group selection*, Blackwell Scientific.
- Zipf, G. K. (1949), *Human behavior and the principle of least effort*, Addison-Wesley.
- Zuidema, W. & de Boer, B. (2009), ‘The evolution of combinatorial phonology’, *Journal of Phonetics* **37**(2), 125–144.
- Zwicker, E. (1961), ‘Subdivision of the audible frequency range into critical bands (Frequenzgruppen)’, *The Journal of the Acoustical Society of America* **33**(2), 248–248.