

Experiments and Simulations Can Inform Evolutionary Theories of the Cultural Evolution of Language

Mónica Tamariz

Abstract Language is a complex adaptive system supported by humans. With evidence and illustrative examples from recent computer simulation and experimental work, this chapter defends that much linguistic structure can be explained as emergent phenomena. Cultural processes such as social transmission to new generations or the patterns of speaker interaction, operating in large populations over many generations give rise to structure at different levels, from categories of phonemes to compositional structure. This position shifts the burden of explanation of linguistic structure from a biologically evolved mental organ to more tractable cultural processes, which are being investigated with a host of innovative empirical methods.

Keywords Cultural evolution of language • Experiments • Iterated learning • Replication • Emergence

1 The Cultural Evolution of Language

The phrase “language evolution” has several meanings. For some, it refers to the *genetic* innovations that appeared in the *Homo sapiens* lineage and that have allowed us to learn, use, produce and understand linguistic behaviour. Sociocognitive and neural capacities such as cooperation, conformity, symbolicity, shared intentionality, imitation or vocal control, which are heavily involved in language, are extraordinarily developed in our species compared to our closest relatives in the phylogenetic tree, namely other apes.

M. Tamariz (✉)

Language Evolution and Computation Research Unit, Department of Linguistics and English Language, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 1AD, UK
e-mail: monica@ling.ed.ac.uk

A second meaning of “language evolution” refers to the *cultural* evolution of linguistic structure. Features of languages like sounds, words or larger constructions can appear, change, move from one language into another, and disappear, giving rise to large-scale patterns of language birth, death and diversification. And all this happens in the historical timescale, through the cultural mechanisms involved in language use and communication in modern humans.

Traditionally, it has been assumed that changes in properties of individual languages such as sound, semantics, morphology or syntax were best explained by *cultural* mechanisms stemming from production and perception biases or population contact. It was concurrently assumed that another kind of properties of languages deemed to be more fundamental, perhaps universal—such as an arbitrary relationship between linguistic signals and meanings, having a closed repertoire of sound categories, coming to be shared by a community of speakers, or being structured in such a way that an open set of novel messages can be produced and understood by other members of the community—required explanations involving *genetic* evolution (e.g. Pinker and Bloom 1990).

One key assumption in this chapter that contrasts sharply with the assumptions outlined above is that cultural processes can explain fundamental properties of language. This perspective has shifted the explanatory emphasis from human genetic evolution towards human cultural transmission. This assumption is the subject of this chapter.

The extended synthesis expands the scope of evolutionary studies both within and beyond biological processes. Thus, the impact of culture on the evolution of humans is the subject of dual inheritance theories, where genetic and cultural information co-evolve (e.g. Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1995; Richerson and Boyd 2005; Odling-Smee et al. 2003). The cultural evolution of language in the second sense proposed above adds a new dimension to evolution: purely cultural processes occur over historical time, sometimes within a few human generations, in a timescale where the biological evolution of humans is irrelevant. New communication systems emerge in the face of novel communicative needs, and those systems evolve through processes indistinguishable from natural selection, neutral evolution, mutation, or gene transfer (of course, applied to cultural, rather than genetic or epigenetic information). The first part of this chapter expounds the assumption that language, a communicative behaviour, is a cultural-evolutionary system (along the lines of Croft (2000), following Hull 1988; and Ritt 2004) and then goes on to describe a selection of recent empirical behavioural experiments and computer simulations whose results are interpreted in terms of that assumption. The second part of the chapter develops the interpretation to lay some foundations for a language-centred theory of language evolution based on previous frameworks [notably Croft (2000) and Ritt (2004)] and informed by the results of the experiments and simulations reviewed in the first part. In this theory, the sounds, words and grammatical constructions we produce undergo replication, variation and selection. Humans are simply a (complex) instrument that mediates replication and selection, while concepts constitute the niches that linguistic items compete for.

1.1 *Language as a Complex Adaptive System*

Many authors investigating cultural evolution have found it useful to consider that language is a complex adaptive system (Gell-Mann 1992; Beckner et al. 2009). A complex adaptive system is composed of many elements that interact with each other. As interactions unfold, their outcomes inform the ongoing interactions. As a result of this self-organization process, *emergent* properties may arise. A classic example of a complex adaptive system is a flock of birds. Each bird has a local rule that attempts to stay within a certain distance and follow the general direction of its close neighbours. The distance and direction can change from one moment to the next, and this change may be affected by the bird's own behaviour. Those are the local interactions. The emergent properties are the flock as a coherent unit and the typical flock motion patterns that look so mesmerizing from a distance. Each bird's individual behaviour is not intended to generate a flock; neither is the flock behaviour predictable in practice from the sum of local interactions—sensitivity to the precise initial conditions is one of the defining characteristics of complex adaptive systems. We know what kinds of patterns to expect, but in a particular instance we cannot predict with certainty what the next state of the flock will be. A good illustration of the notorious difficulty to predict the behaviour of a CAS is the weather, with countless air molecules interacting under local conditions of pressure and temperature from which the likes of storms, tornadoes or spells of dead calm emerge.

In language, we have a *local level* where individual instances of linguistic behaviour are produced typically in interactions between speakers for particular, usually communicative, purposes in a given context. A large number of such interactions give rise to *emergent properties* such as linguistic regularities and categories or coordination of conventions at the level of the population. The actors in a particular linguistic interaction normally do not intend to generate a population-wide system; rather, they just want to communicate about something, there and then. Any population-wide or language-wide patterns are unintended side effects, emergent properties or, in the terminology used by Keller (1994), results of the action of the “invisible hand”. Moreover, predicting the long-term outcomes of language change at the population level is so complex that it has not even been attempted.

1.2 *Evolutionary Processes*

Over the course of this chapter, we will see evidence that cultural transmission can lead to self-organization and emergence, but also that it involves evolutionary processes such as inheritance, variation, neutral evolution and selection. The first part of this chapter discusses recent studies that focus on different aspects of language evolution. The model of cultural language evolution presented in the second part puts this discussion in the context of other frameworks of the cultural evolution of language and, crucially, relates processes to mechanisms, for instance, inheritance

to babbling and language learning; variation to social mechanisms of generation and spread of variation in the sociolinguistics literature; and selection to social interactions and to the structure of meanings.

Languages do change, but they are also remarkably stable. Studies of language families (Dunn et al. 2005) and individual words (Pagel 2009) claim to have reconstructed lineages of language traits up to 10,000 years into the past. This stability is due to linguistic information being culturally inherited by new speakers, in other words, by infants very faithfully learning the language of their social group. This is why the mechanisms of language learning by infants are of interest to the study of language evolution. But we do not stop learning when childhood ends. Rather, learning continues throughout speakers' lives, as we are exposed to and create innovations during usage. Therefore, the mechanisms of language usage by adults are also of interest to cultural language evolution.

The low-level mechanisms that mediate the inheritance of linguistic information, both during learning and over usage, include imitation and conformity. An important aspect of imitation is copying behaviour irrespective of its function, or even of whether it has a function. This human capacity typically enables the accumulation of traits that leads to the cumulative complexification (Heyes 2013; Boyd and Richerson 1995; Tomasello 1999) of language structure.

Inheritance is not perfect, however, or there would be no evolution. Innovations can be introduced into languages through production and perception errors, as the result of speakers' efforts to express novel meanings, or through contact with other languages. And once we have different variants of the same linguistic items (a sound, a word, etc.), there may be evolutionary competition between them.

Neutral evolution, defined for biology (Kimura 1983), where variants spread following random dynamics, has been highlighted as an important mechanism of evolution in language (Nettle 1999) and culture (Bentley et al. 2004; Herzog et al. 2004). But selective pressures are also at work and a host of factors can affect the structure of languages as they are used and transmitted to new members of the population. Some examples of selective pressures are as follows:

- Cognitive biases mean that certain words, sounds or constructions are more likely to be used because they are easier to learn, process, produce or perceive than others. The preferred variants may end up being more frequent in the language than their competitors.
- Social constraints that make speakers prefer linguistic elements that are original, fashionable, conformist or complex and features that serve to identify speakers as part of a group, or to distinguish them from another group will also leave their mark at the language level.
- Patterns of connectivity in the population, influencing how many other speakers one interacts with and how often, or whether the patterns are homogeneous in the whole population or not can have an impact on the structure of emergent languages.
- The structure of the world and particularly any features of the meanings that speakers want to communicate about can also affect the ways those meanings are expressed.

These selection pressures may result in certain linguistic patterns being more likely to be learned by new individuals, more apt to spread in a population, or more efficient for the purposes of communication. The following section describes recent computer and behavioural models of the evolution of communicative systems, which are then analysed in terms of the evolutionary processes they reveal.

2 Computer Simulations and Experiments

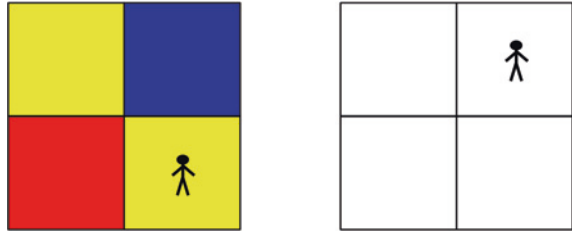
Aside from having important theoretical implications, prioritizing cultural-evolutionary explanations over biological-evolutionary ones has opened new methodological avenues to explore the origins of linguistic structure, especially with experiments that model the transmission of linguistic information both during communicative usage and over learning by new generations of speakers. Mathematical and computer simulations have been applied, hand in hand with behavioural experiments, in many successful lines of research: experiments can verify simulation results, and simulations can be used to construct models based on experimental results. This section specifically reviews and discusses a selection of experiments and associated computer simulations looking at how cultural-evolutionary processes shape the structure of languages.

The following sections describe a selection of studies that explore the creation of conventions, or individual signals that have an agreed meaning for the interlocutors that use them; the spread of conventions through a population; the emergence of cultural systems such as vowel systems; and finally the cultural emergence of linguistic structure. I do not present an exhaustive literature review, but rather a sample of classic and new experiments and simulations with the aim of illustrating the evolutionary way of thinking about the language transmission and usage sketched above. Each study highlights an aspect of this approach and will be accompanied by a discussion of the methods and results in terms of evolutionary processes and elements. Along the way, I will point out further questions that could be tested empirically by extending or adapting the studies described. Finally, the last section brings those elements together to outline a theory of the cultural evolution of language.

2.1 The Emergence of Conventions

Imagine you are on holiday in a country where you have no common language with the locals. One morning, you go to reception to borrow a hair dryer. You put to work your best gesturing abilities to describe what you want, and the reception man tries to be as helpful as possible. In the midst of your gesturing, you say “Wet hair! Wet hair!” and at that precise moment, your interlocutor produces a hair dryer from a drawer. Upon seeing your happy face, the man says with a relieved

Fig. 1 One player's view of the game: *on the left*, the player's own board; *on the right*, the partner's board. Adapted from Scott-Phillips et al. (2009)



understanding smile “Wet-air! Wet-air!” The next morning, when you need the hair dryer, you ask the man for “wet-air” and immediately get what you want. The new word may be used by your friend when she wants to borrow the same item and you advise her to ask for “wet air”. Similarly, the other receptionist may overhear and learn the new word and thus be able to help you or your friend with your requests in the future. You have created a new convention, and it has begun to spread beyond its creators, namely you and the receptionist.

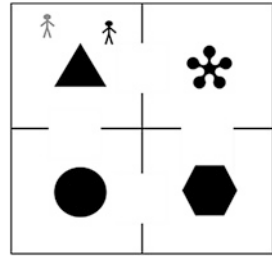
Several things need to be in place for this to be able to happen. You need to have a shared communication channel, in this case gesture, that both you and the man at reception interpret as such signals.

For this to happen, the receptionist has to recognize your gesturing as communicative. This seems such obvious parts of the communicative interactions we are involved in every day that it is easy for us to take it for granted. However, it is not as obvious as it seems. Chimpanzees in the wild, for instance, do not interpret signals such as pointing as communicative without training, and tend to look at a pointing finger instead of at the direction where it is pointing (Povinelli et al. 2003). Humans, in contrast, have the inclination to interpret signals as communicative and also have a dedicated channel, most commonly speech, through which we expect to receive communicative signals. Even if one of these conditions is not met, we can still get by with the others, as in the hair dryer example, where there was not a shared language, but the gestures were readily understood as communicative.

But if there were no dedicated medium, shared signals or pre-established signaller and receiver roles, would we still be able to create an effective communication system? Two studies by Scott-Phillips et al. (2009) and Galantucci (2005) address this precise question. Scott-Phillips and colleagues designed an experiment where two people sat at two connected computer terminals and played a cooperative game. Each player had a playing board, composed of four coloured quadrants (Fig. 1) and a little character that could be moved from the centre of one quadrant to the centre of an adjacent quadrant by pressing the arrow keys. The quadrants could be red, green, yellow or green. The aim of the game was for both players to place their characters, each on their board, on the same colour.

Each player saw on the screen both their own and their partner's playing boards and characters, but they could not see the colours in the partner's board. The players could not see or hear each other, they were simply told about the task. In

Fig. 2 Galantucci's game map: four rooms connected by doors. Adapted from Galantucci (2005)



these conditions, agreeing on a colour seems impossible—can you come up with a solution? After playing some games, however, most pairs managed to create a communication system, and in all cases, the process was the same. The first step to success consisted on the two players implicitly deciding on a default colour, so they would both land on that colour and thus score a point. The default colour tended to be red, perhaps the most salient one. But this strategy had a problem: sometimes, red was not present in one or both boards. The next step to success happened as a reaction to this situation. When a player did not have red on her board, she would attempt to let her partner know by moving her character for instance from left to right and back to the left, repeatedly. Then, she would land on another colour, say blue. This situation prompted the association of blue with left–right motion. At this point, the first communicative signal was created, at the same time as a communication medium, the character's motions, was discovered. After this insight, the other two colours were soon associated with other character motion patterns, and the game task, supported by communication, became trivial.

This experiment highlights the versatility of communication systems and the resources of our drive to communicate: any information pattern can become a signal. The characters' movements, which in principle served simply to move between quadrants, are co-opted for a communicative function. Also, the whole system is bootstrapped from an initial heuristic, namely always landing on the same colour and a situation where the heuristic fails.

A somewhat related study addressing the creation of a communication system de novo is that of Galantucci's. Here, two people also played a cooperative computer game; in this case, the task was for the two characters to meet in the same quadrant; quadrants were now identified not by colour, but by a shape (Fig. 2). During the game, each player only saw the room he or she was in and had to infer the whole map from experiencing moving through the doors.

There was an additional important difference between this experiment and the one described above: here, there was a dedicated communication channel. Beside each computer were a digital pad and pen, and the players could see on their screen what they and their partner wrote or drew. The experiment instructions did not mention this pad, and if the players attempted to use it, they would discover that what they wrote was heavily distorted, as if they were writing on a moving tape and faded rapidly from view.

As before, it is difficult to score points consistently in this game without communication and, here again, not all pairs of players found a solution. Those who succeeded, however, developed signals that were adapted to the communication channel and to the meaning structure. The distorting-pad prevented players from writing letters or numbers, or from drawing the figures found in the rooms; the signals they employed were those most immune to the distorting effects of the pad like vertical lines or small marks. As for the meanings employed to identify the rooms, several strategies were apparent. Some pairs numbered the rooms by drawing one, two, three or four small marks on the pad. Others tried to represent the triangle, circle, etc. iconically, for instance by relating to the number of vertices in each figure: three for the triangle, five for the star, one for the circle and six for the hexagon. A third group drew a line on either side of the pad to indicate whether they were in a room on the left or on the right. The last solution is ambiguous, since there are two signals for four different rooms, but it was complemented by the following strategy: at the beginning of a game, each player would tell the other which side of the board they were in. Then, one of the players would always move first. This combination of ambiguous signal plus conventionalized turn-taking allowed success at each game.

In two continuations of this experiment, Galantucci progressively increased the task's difficulty. The successful pairs from experiment 1 went on to the next level, which had a 3×3 board, and then to a 4×4 one. The drawings in the new rooms included an umbrella, a bird, a star, a hash or a crown. Performance levels in the complex environments turned out to be dependent on the type of strategy employed at the easier level: numerical systems were easily adapted to the extended boards, but iconic and side-based systems were not. The constraints of the drawing pad did not allow to keep drawing new sufficiently distinct icons for the figures. And the ambiguity of the side-based system grew exponentially with each increase in board size.

These differences can be understood in terms of adaptiveness of the signal systems to the game tasks. The main function of a communicative signal is unambiguously to point to one meaning among several possible ones (in this case, one room out of the four, nine or sixteen in the board). One of the traits that make signals useful, or likely to be reused, and therefore "fit", is being distinct from each other. In the initial, 2×2 board, the iconic, number-based and position-based-plus-turn-taking solutions were all adaptive. But in an extending meaning space, only the numerical one proved to be adaptive. In the former case, we can talk of a number of adaptive *independent* signals, each distinctly pointing to one room. However, in the latter, it is the *system* that is adaptive, as it allows for the repertoire of signals to be extended in a way that preserves and expands the communicative function. We will see more of adaptiveness at the system level when we talk about experiments with languages below.

In a further study dealing with the early evolution of signals (Garrod et al. 2007), pairs of participants created graphical conventions to represent a series of concepts in a pictorial-like task. In each game, the "director" and the "guesser" each saw a list of sixteen concepts. One of these concepts was selected and given

to the game director, who would draw something for the guesser. If the latter guessed correctly what the target concept was, the pair scored a point. Garrod and colleagues were specifically interested in the role of feedback between the players in the emergence of arbitrary signals, and their experiments involved two kinds of feedback: one manipulated whether the two participants exchanged the roles of drawer and guesser in the games and the other whether the guesser could interact with the drawer, for example by asking for clarification.

Their results revealed that the players obtained higher scores when both types of feedback were allowed. Moreover, the drawings became simpler and more arbitrary (less likely to be identified by onlookers) only if at least one kind of feedback was available.

Brown (2012) has criticized the conclusion that the final simple drawings in these experiments are truly arbitrary, arguing that the even if onlookers could not identify the drawers accurately, the players themselves could still see traces of the initial iconic relationship between drawing and concept. For instance, two inverted V's would be impossible for an onlooker to recognize as a representation of the concept "cartoon". But the players and creators of the signal would probably still interpret the V's as stylized versions of a cartoon rabbit's ears. However, after a few generational transmission events, new users of the signal would not know about its origin and, therefore, for them, it would be truly arbitrary.

In Garrod et al.'s (2007) experiments, two-way interaction between the players allowed them to know that they were "on the same wavelength": they knew what a drawing meant for both of them, in the context of the game (e.g. a rabbit had been enough to make the guesser select "cartoon" from the sixteen possible concepts). They knew how and why the drawings changed over the course of the game (e.g. one of them drew the rabbit's ears and, before he had time to finish the drawing, the guesser gave an answer; from then on, the ears would be enough to identify the concept "cartoon"). This neatly demonstrates how a purely cultural process, social interaction, can contribute to the emergence of arbitrary communicative signals.

A cultural-evolutionary analysis of the three studies described can be framed in terms of adaptation of the emerging systems to constraints. A pressure inherent to these communicative tasks, indeed, the function of any communicative system, is *expressivity*: if we are to avoid ambiguity, a distinct signal is required for each meaning. The expressivity of a system depends in turn on the flexibility of the *communication medium*, which is decreasingly constrained in the three experiments. The main difficulty in Scott-Phillips et al. (2009) experiment was that the players needed to *exapt*, or co-opt, the characters' movements, whose original function was simply to move from one quadrant to another, for the novel function of providing a communication medium. Galantucci's (2005) and Garrod et al.'s (2007) experiments had a dedicated graphical communication channel, but in the case of the former, rapid-fading and linear motion prevented drawing normally. Nevertheless, some simple signals arose that were easy to produce within these limitations and sufficiently distinct from each other. As the expressivity pressure increased as new rooms were added to the board, the contrived communication medium made it impossible for some of the emergent systems to meet the

expressivity requirements, and the participants failed to solve the task. Garrod et al.'s participants needed to disambiguate among sixteen referents, the same number as in Galantucci's third experiment, but in this case, they used drawing, a familiar medium both for game directors and guessers. This made all the difference, and all players succeeded in the communicative task, which suggests that given more time and practice Galantucci's participants would have the opportunity to explore the vast space of possibilities offered by their limited communication channels and evolve complex, structured systems. I say vast because, after all, rapid-fading and linearity are also characteristics of speech, which is short-lived and does not allow going back to revise what was produced earlier.

2.2 The Spread of Conventions

The players in Scott-Phillips et al. (2009), Galantucci (2005) and Garrod et al. (2007) managed to create new communicative conventions and to use them successfully over and over to achieve a goal. In real languages, innovations like new words and expressions are created all the time, usually spurred by the context, by knowledge shared by the interlocutors, by the need to express a new meaning, or by the desire to express something in a new way. Most of these innovations are never used again or, as in the games above, are only ever used by their creators, but some may catch on among new interlocutors of the creators, like the intercultural solution "wet air" for hair dryer in the story spread to a friend and to other people working in the hotel. A subset of all innovations will be adopted by more and more speakers and a few among them may spread to a whole linguistic community via the connections of the social network.

Coordination, or convergence on the same solution to express a meaning by the whole population can be viewed as an emergent property of languages. At each interaction, the interlocutors just wish to communicate with one another; but complex patterns of interactions involving many interlocutors may result in coordination at the population level. This hypothesis that shared conventions emerge through cultural processes of self-organization was tested in series of influential computer simulation studies carried out by Steels (1996, 1998, 2003, 2006).

The basic skeleton of these simulations, laid out in Steels (1996), includes a population of agents who are able to learn associations between signals and meanings, and who interact in pairs playing "language games". In each game, a speaker and a hearer are selected from the population and a set of objects are chosen as the context of the game. One of these objects is marked as the topic of the game. The speaker names a distinguishing feature of the topic (e.g. the colour and shape, that singles it out from the rest of the context), and subsequently the hearer points at the object that he thinks the speaker was referring to. With this information, both agents update their vocabulary in order to align to each other. Over many such interactions, involving many different player pairs, the agents' vocabularies,

initially empty and subsequently idiosyncratic to each pair, end up being shared by the whole population.

This dynamics is mirrored in another experiment where eight people play pictorial games in pairs (Fay et al. 2010). The pairs change several times so that in the end, every player has played with everyone else. Within each pair, the game procedure was the same as in Garrod et al. (2007) where there was role swap and feedback. The initial pairs of players normally developed different representations for each concept, for example, the concept “cartoon” could be represented as the drawing of a rabbit for a pair, a Simpson character for another and Mickey Mouse ears for yet another. But often, after several partner changes, the whole population converged onto the same representations. In some cases, however, two or more variant representations remained.

How did this happen? The patterns of variant spread observed in Fay et al.’s population could have resulted from neutral evolutionary dynamics whereby variants that happen to have a higher initial frequency have a higher probability of spreading to the whole population through random processes. (In biology, this dynamics is instantiated as genetic drift.) In fact, Steels characterizes the spread of information in the population in his 1996 simulation described above and in other associated studies as neutral evolution, where all the possible ways to name the objects are equally likely to spread. Several studies have highlighted the power of neutral evolution to explain cultural evolution, for example, the names given to babies (Bentley et al. 2004) and the breeds of dog that people tended to buy (Herzog et al. 2004) changed according to the neutral model of evolution. Others have focused on the role of selection pressures on the spread of cultural variants (Richerson and Boyd 2005). Linguistic innovations are believed to spread by a mixture of neutral and biased transmission (Nettle 1999; Trudgill 2004; Blythe and Croft 2009).

The nature of variant spread in Fay et al.’s (2010) experiment has been explored in a recent simulation study (Tamariz et al. in preparation). Mirroring the experimental design, the simulations had eight agents arranged into pairs playing communication games about sixteen concepts, and undergoing the same number of interactions and partner exchanges as in the experiment. The spread patterns obtained in the simulations were compared to those recorded in the experiment.

The study considered several possible mechanisms of spread: first, a model of neutral evolution where each agent chose which variant to produce randomly from the two variants he saw in the previous round—his own and his previous partner’s. Some empirical data points, for instance those for concepts where convergence had not been achieved, were similar to the results of the neutral evolution simulations, but the majority converged faster than predicted by this model. The second model included a bias for conformity that increased the chances that the two players in a pair would coordinate by agreeing on the same representation variant. Some of the remaining data points were captured by this model, indicating that local coordination could also be at work. Indeed, the goal of each interaction is to communicate successfully with the current partner, and using the same variant as him or her is a good strategy to achieve this. The third and final

model tested included a selective pressure, namely content bias: certain variants were intrinsically more likely to be copied than others. This model captured all the remaining data points, and in fact the models that best fitted most empirical spread patterns included a degree of content bias.

These results further support that multiple processes (neutral evolution, content biases, social heuristics) may operate on the evolution of culturally transmitted variants. Additionally, they provide evidence for replicator dynamics operating at the cultural level (Blythe and Croft 2009): the computer models in Tamariz et al. (in preparation), assumed that each representation variant was, in general, faithfully replicated (when it was reproduced by a player at each round in the game) but could also mutate (when it was modified by a player).

This study reveals how some variants spread while others do not, but it does not address *what* makes some variants more likely to be copied than others. Are the winning variants easier to produce, process or perceive than others? Are they particularly clever or ingenious? Are they simple? Or iconic? Future manipulations of the type of cultural information transmitted in experiments like Fay et al. (2010) informed by studies on cognitive salience combined with computer simulations such as those in Tamariz et al. (in preparation) could help us understand precisely what kinds of content biases operate on cultural transmission.

2.3 *The Cultural Emergence of Systems*

In the work reviewed thus far, new conventions emerged and spread in a population. Participants in different experimental games started to use their own behaviours as signals, with communicative intent or to designate concepts; but these signals were largely independent from each other, there were no categories, rules or interactions that justified treating them as a unified system. This contrasts with many cultural realms, notably language, which behave, as we have seen, as complex adaptive systems. Next, I describe some studies where properties typical of systems emerge out of cultural processes of interaction and transmission, and then, I discuss the evolutionary processes they reveal.

De Boer (1999, 2000, 2001) observed that languages have a fixed repertoire of vowels whose organization is strongly constrained. In particular, vowel categories are widely spread out in the acoustic-articulatory space so they maximize occupation of the available space and minimize overlapping. This results in vowels that are easiest to distinguish from each other by hearers. de Boer designed a simulation with two agents who could produce, perceive, and remember speech sounds characterized by three realistic parameters: tongue position, height and lip rounding. The two agents played repeated imitation games (swapping roles at every interaction) as follows: the initiator would select a vowel prototype from memory and produce a version of it (with noise). The imitator would hear it, interpret it as one of the prototypes in his memory, and produce a version of this prototype (again, with noise). The initiator would then interpret this as a prototype; if this

was the same as the initial prototype, the game was successful. Both agents could then update their memories with the information from the game by merging close vowels, throwing away seldom used vowels, adding new random vowels or moving produced vowels closer to the perceived ones within a prototype. After many iterations of the game, the vowel categories spread over the whole articulatory space and maximized the distance between them, optimizing distinguishability like natural vowel systems. This final arrangement is an emergent property at the level of the whole system.

Wedel (2006) offers an interesting evolutionary approach based on exemplar models to explain the mechanisms for the formation and stabilization of vowel (or other sound) systems such as those modelled in de Boer (2001). Each sound category exists as a distribution of “exemplars”, or individual instances of production in a linguistic community. We learn by hearing and producing many such sound exemplars, which all leave a memory trace in our minds. Each individual’s knowledge of a sound is based on a different sample and therefore is slightly different from every other speaker’s knowledge. With such pervasive individual differences, how come sounds do not change constantly over time and across a population? Wedel suggests the answer is *blending inheritance*, a mechanism for the transmission of features with continuous values that is capable of maintaining stable categories over time. Each exemplar of, say, vowel *oo*, we produce is based not on a single exemplar in our memory, but on all the exemplars in that vowel category. The production target may have the average tongue position, height and lip rounding values of all the *oo* exemplars in our memory. It is harder to argue for replication in this case, as each production is clearly not necessarily causally linked or similar to a particular previous exemplar, and establishing lineages is therefore not obvious. Nevertheless, Wedel persuasively argues that the categories behave as discrete replicators. The stable categories achieved with blending inheritance translate into a multimodal distribution of acoustic values in the output data produced by speakers. New learners exposed to this input will go on to produce (by blending inheritance) sounds with the same underlying distribution properties as the one they learned, so, over the generations, both the statistical properties of the distributions of exemplars produced and the categories are maintained. As Pierrehumbert puts it, “phonological representation (is) an error-correcting code” (2012: 175).

Recently, Verhoef and colleagues devised a behavioural experiment somewhat related to de Boer’s vowel category simulations, modelling the emergence of a system of signals (Verhoef et al. 2011, 2012). In this case, the cultural mechanism at work was not interaction in a pair (*closed-group method*) in Mesoudi and Whiten’s (2008) terminology, but transmission to new learners (*transmission chain method*). Verhoef and colleagues had a human participant learn how to play a set of twelve different, random whistles with a slide whistle—a kind of flute with an embolus inside that can be pushed in or out to produce higher or lower sounds. He or she would subsequently attempt to play back the twelve sounds, no repetitions allowed. These twelve new whistles were used to train the next participant in the diffusion-chain experiment (see next section for more details on this paradigm).

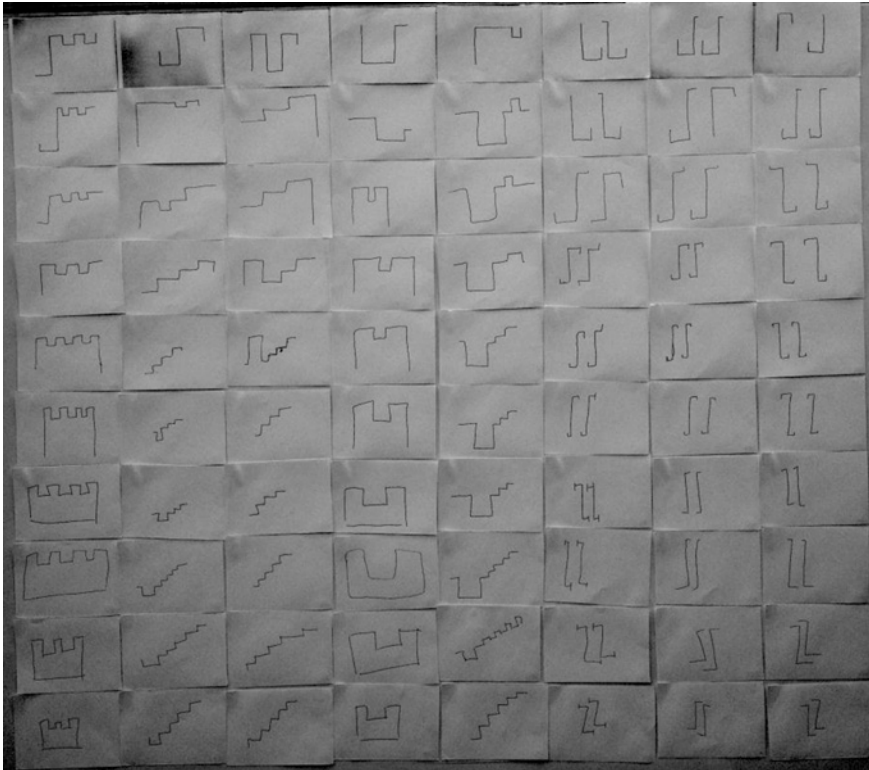


Fig. 3 A transmission chain of drawings involving nine participants. The *top row* shows the eight original drawings, and *subsequent rows*, the drawings produced by each of the participants

The second participant's output would be the input to the next, and so on for ten "generations". Over repeated transmission, the sounds changed dramatically and the final whistles were typically a *system* composed of a small number of sub-whistles recombined in different ways. At each episode of transmission, the structure of the information transmitted changed a little, but this happened without any intentionality on the part of the participants, who were merely trying to reproduce what they had learned. Over the generations, some of the more difficult-to-remember patterns would not be reproduced in the output, while novel whistles would be produced; the pressure to produce twelve different sounds encouraged reuse of remembered sub-parts in those novel whistles produced. In addition, the sub-parts were sometimes reduplicated or reversed. The final systems were, consequently, simpler in the information-theory sense—they had much lower entropy, i.e. were more compressible than the initial ones—and easier to learn—later-generation participants only needed to memorize a few units and a few ways to recombine them. Out of the initially continuous sound space, a small set of discrete patterns emerged. These are usually very different from each other, as was the case with

the vowel categories that emerged in de Boer's simulations and therefore easy to distinguish, memorize and produce.

Figure 3 shows some results from a recent small study carried out by the author of this chapter as a teaching exercise. The design of this study was the same as that in Verhoef et al.'s (2011, 2012), but here, eight drawings were transmitted instead of twelve whistles. The ten generations of drawings show visually the same kind of recombination and simplification processes attested in the results of Verhoef et al. (2011, 2012). The processes involved here, again, were reduplication reversal and recombination. For example, the sixth participant in this chain (7th row of drawings in Fig. 3) reversed the sixth drawing (or perhaps recombined the shape of the rightmost drawing and the lines at the end of the Z-shapes from the third drawing from the right at the previous generation) and also added a horizontal line at the bottom of drawings one and four. It is also apparent that three categories of drawings become increasingly obvious over the generations: the three Z-like drawings on the right; the three drawings with steps (drawings two, three and five) and the two closed drawings (one and four). A short description of the final set of drawings, then, would probably involve the three categories plus the details that differentiate the items within each category, for instance the direction of the end lines of the "stairs" or the direction and presence of endlines in the Z-like drawings.

The drawings, initially unrelated to each other, evolve to form a *system* of related categories. I argue that the emergent units (the categories and the details) are replicators, as they fulfil the necessary criteria of similarity (copying fidelity), causality and information transfer between model and copy, longevity and fecundity (Dawkins 1976; Ritt 2004; Sperber 2000; Godfrey-Smith 2000). Once the sub-units in the whistles in Verhoef et al. (2011, 2012) and in the small drawing experiments have stabilized, after a few generations, they are faithfully and reliably copied in such a way that we can trace their lineages; they are causally connected to previous productions of the same units; and information about their structure is transferred from originals to copies.

Additionally, the small number and systematic nature of the replicator set at the final generations is an adaptation to the elements involved in replication. First, limited exposure time to the original whistles or drawings only provides limited opportunity to memorize long complex patterns. Second, a cognitive preference for regularity leads the participants to reuse patterns and processes they have memorized and extend them to the whistles and drawings they cannot remember well.

2.4 The Evolution of Regular Linguistic Structure: Systematicity Between Signals and Meanings

So far we have seen evidence that individual *communicative signals* on the one hand and structured *systems* of replicable units such as whistles and drawings on the other can emerge and stabilize over repeated use and transmission. Next, we will look at simulations and experiments showing that structured *systems of*

signals, such as artificial languages, can also emerge from the same cultural processes.

All natural languages are compositional: the meaning of a complex linguistic utterance is a function of the meanings of its elements and the ways in which they are put together. For instance, the meaning of a sentence such as “man bites dog” depends on the meaning of its components and their order (compare “dog bites man”). Compositionality is thus the key property of languages that allows us to recombine words and constructions in infinite ways to express new meanings. Kirby (2001) published a simulation study that demonstrated that compositionality can emerge from cultural transmission dynamics alone. This was the first of an ongoing family of simulations and experiments applying the iterated learning model to explore the role of transmission in language. Iterated learning is the “process in which the behaviour of one individual is the product of observation of similar behaviour in another individual who acquired the behaviour in the same way” (Scott-Phillips and Kirby 2010: 411). Kirby’s (2001) seminal simulation involved chains of agents learning artificial languages composed of a set of meanings, and their names. Every generation in the chain included three steps: first, an “adult” agent was given some meanings and it must name them using the signals in its memory, or inventing; next, a learner agent learns the language (the associations between meanings and signals) produced by the adult; finally, the learner becomes the adult for the next generation and the old adult is discarded. Learning involved agents adding the new associations to their memory and streamlining redundant information; here is an illustration of the streamlining process: if an agent had the associations {john, eats} \Rightarrow “johneats” and {tiger, eats} \Rightarrow “tigereats”, the streamlining process will replace them with the more general {*x*, eats} \Rightarrow “*xeats*”, {john} \Rightarrow “john” and {tiger} \Rightarrow “tiger”.

The initial signals were invented random letter strings produced by the first adults, but the languages kept changing over the generations until eventually they stabilized. The crucial result was that the final, stable languages resulting from the simple iterated learning dynamics were compositional: each value of each meaning dimension had an associated letter string; and letter strings were combined according to rules to form the complex meanings. Kirby’s insight was that the language adapted to the transmission process: unlike the initial ones, the final, compositional languages were stable and did not change over transmission. Even if the input only included part of the language, its compositional structure allowed the agents to reconstruct the complete language. If compositionality, a fundamental property of languages, can be explained by cultural processes of learning and transmission, reasoned Kirby, perhaps we should focus more on how languages have adapted to humans through cultural mechanisms and less on how humans have adapted to language through biological-cognitive evolution.

This and related computer models inspired a line of experimental work that confirmed and complemented the simulation results. The first modern experiments on artificial language iterated learning are those described in Kirby et al. (2008). The dynamics closely mirrored those in Kirby (2001): here, a human participant was trained on (a half of) an artificial miniature language: words referring

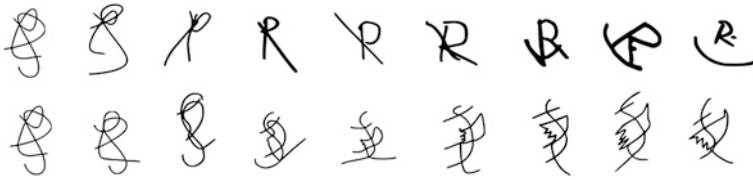


Fig. 4 Drawings from representative chains in the two conditions in Tamariz and Kirby (in press). Generations 0, 1, 4, 7, 10, 13, 16, 19 and 22 are shown from a chain in the memory condition (*top*) and one in the copy-from-view condition (*bottom*)

to objects; then, she was asked to name the full set of meanings. From her output, half of the object–word pairs were selected and given to the next learner as training input. The learner then was asked to name *all* the objects, and so forth for ten generations. Each language consisted of 27 words referring to as many meanings: all the possible objects combining three shapes (triangle, circle and square), three colours (blue, black and red) and three motions (spiral, horizontal and bounce). While in the simulation the first agent (generation) in a chain invented its own signals, in the experiment the first participant was given randomly constructed words. The results revealed that the initial languages where each object was associated to a random word became structured. For instance, in one transmission chain, all the objects moving horizontally ended up being called “tuge” and all the objects with a spiral motion, “poi”.

The languages in this experiment and in Kirby’s (2001) simulations became increasingly stable over generations. This can be interpreted, again, as the emergence of stable replicators. In the absence of any communicative requirements, the only task given to participants was to learn and reproduce the language as faithfully as possible, and the languages readily adapted to the task. Moreover, the fact that the word-replicators referred to categories of objects (those moving horizontally or spirally) indicates adaptation of the words to the structure of the meanings. The few remaining words in the final language were associated not to random collections of objects, but to meaningful categories.

These results were different from those in Kirby’s (2001) simulation in an important respect: there were a lot fewer words than meanings. The simulations had an implicit bias for unambiguous mappings where a distinct signal was associated to each meaning. In a second experiment, Kirby et al. (2008) also introduced a bias for diversity, or expressivity when they had to select the items from one participant’s output to construct the training set for the next one, instead of doing it randomly as in the first experiment, they removed as many items with duplicate words as possible. The languages emerging in this condition did not stabilize to the same degree as in the first experiment, but reproduction fidelity kept increasing. But the most dramatic effect of this subtle manipulation was the emergence, in some languages, of *compositional* structure over the generations. In the languages where compositionality emerged, different parts of the words (the beginnings, middles and ends) became associated with colour, shape and motion

categories. In one of the chains, whose final-generation language is represented in Fig. 4, word beginnings were reliably associated with colour, and word-endings with motion. In a perfectly compositional system, word-middles would be associated with shape; and there would be no variability such as the one found in the example in Fig. 4 for colour blue (which is expressed as either *ku* or *hu*). This variability, incidentally, does not introduce ambiguity for comprehension, but it creates a degree of uncertainty in production—should I use *hu* or *ku*?

In Kirby et al.'s (2008) second experiment, replicators also emerged. In this case, they were not whole signals like the “tuge” or “poi” words from the first experiment, but elements in the signals, and their positions, which also became increasingly stable over the generations—in other words, evolved to become easy to replicate.

In contrast with the signal-only systems (whistles, drawings), in systems where signals are paired with meaning, certain letter strings (words or parts of words) become increasingly consistently associated to features of the drawings and thus acquire a symbolic or referential function. The “meaningful” strings, in turn, become easier to remember—to replicate—both because they “mean” something and because of their higher frequency. When a participant needed to name an object that she had not seen during training, say, a red horizontally moving triangle,¹ she would be likely to use letter strings that had been associated with red colour, horizontal motion or triangular shapes in her training set. The adaptive solution that emerged over the generations under the double constraint to be easy to replicate and functional was a compositional system: only nine segments, combined in different ways, could express 27 meanings. Compositionality is, in fact, an efficient solution to the problem faced by languages unambiguously naming a structured set of items (Brighton et al. 2005). Compositional languages have three important interrelated properties: first, they are *expressive*, as they allow one-to-one unambiguous mappings between signals and meanings, which are good for communication. Second, they are *compressible*, as many meanings can be expressed with few signals; this increases the replicability of the system, since only a few items have to be memorized. And third, as a consequence of the first two properties, the languages are extendable; in other words, they allow the expression and understanding of novel meanings.

The difference between experiments one and two that allowed the emergence of compositionality only in the second was the removal duplicate signals from the training input. In a real language, this bias should arise from the need of words to refer to meanings unambiguously if the communicative interaction is to be successful. This last point was tested in a recent series of experiments, where Kirby et al. (in preparation) combined the iterated artificial language learning paradigm with communicative tasks. The main question they addressed was: what is the relative contribution of transmission to new learners (transmission chains, as in Kirby 2001; Verhoef et al. 2011, 2012; Kirby et al. 2008) and usage [closed-group

¹ For the most part, participants were unaware that there were asked to name novel objects.

interactions, as in the communicative experiments by Scott-Phillips et al. (2009), Galantucci (2005), Garrod et al. (2007) or Fay et al. (2010)] to the emergence of structured systems? In their experiments, Kirby et al. (in preparation) had pairs of participants play a game on two computers. Following Kirby et al. (2008), they were first trained on a language; but then, instead of simply reproducing the words, they *used* the language to play a cooperative communicative game together. The last words produced for each of the objects at the end of their game were used to train the participant pair in the next round in the game, and so on for six rounds. The key manipulation was whether the next round was played by the same pair of players (closed group) or a new pair (transmission chain).

The communicative interactions between the pairs were inspired by the “language games” in Steels’ (1996) computer models and involved a speaker and a guesser. The speaker was shown a target meaning, which she had to name. The guesser then saw on his screen the signal typed by the speaker and an array of possible meanings, and had to choose which of the meanings corresponded to the signal. If the chosen meaning was the same as the target, they jointly scored a point and the players exchanged roles at every game.

Overall, the final languages were, as in previous experiments, easier to learn (more stable across rounds) and more structured (systematic and compositional) than earlier ones. As far as the role of the communicative task is concerned, the hypothesis was held: the final languages contained hardly any homonyms, so they were fully expressive and effective for communication (more points were scored at the final than earlier rounds). In other words, the communicative task achieved the same effect as the expressivity pressure introduced in experiment 2 in Kirby et al. (2008). So, as long as we are not interested in other aspects of communicative interaction [such as feedback, as in Garrod et al.’s (2007) pictorial studies], the simple iterated learning-and-recall with anti-homonymy filter is a valid design. If, however, we wish to explore aspects of communicative interaction, we can choose the iterated *communication* design.

As for the roles of closed group versus transmission chains, the languages in the closed-group condition—where the same two people went through repeated rounds of training and play—obtained higher scores and were more stable than those in the chain conditions, where the pair was replaced by a fresh one at each round. Crucially, the chain languages showed a strong increase in compositionality, but those in the horizontal condition did not change in that respect and remained a set of distinct, but idiosyncratic and unrelated signals.

It seems, then, that transmission to *new learners* is the key element for structure to emerge in these experiments. In evolutionary terms, they represent the generation turnover that renovates the pool of linguistics variants. New learners bring to the dynamics the necessity for inheritance of information between generations, a crucial element in any Darwinian system. In the iterated communication experiments, having several consecutive learners exerts pressure for stable, faithfully replicable languages; and communication selects for expressive languages. The adaptive solution to this double constraint is, again, compositionality: a simple system with few elements to memorize and reproduce, and a few combinatorial

rules that make the languages robust against memory failure, because it is possible to generalize from a few items to the whole set of meanings.




For compositionality to be at all possible, an additional requirement is that the meanings are structured. The repetition of meaning features (colour, motion and shape) in different objects requires that several meanings share the same features. I have spoken of this match between words and object categories in terms of adaptation of forms to meanings. This hypothesis is upheld by an experiment by Perfors and Navarro (2011) that shows the impact of the structure of the meaning space on the final language. They run iterated language chains like those of Kirby et al. (2008), but used an especially designed set of meanings: squares of six different sizes and six different levels of brightness.

They manipulated the structure of the meaning space to make one dimension more salient than the other. In the control condition, the values of size and brightness were evenly distributed. In the *size condition*, there were three smaller sizes and three clearly larger sizes. In the *brightness condition*, there were three lighter and three clearly darker shades.

Their words were consonant–vowel–consonant syllables, and they did not apply a filter to eliminate homonyms, so they expected ambiguous languages to emerge, similar to those of Kirby et al. (2008, experiment 1). However, while Kirby et al. had no prediction as to which meaning would come to be expressed (e.g. motion in the case above, or colour, or shape), Perfors and Navarro's (2011) manipulation of the meanings predicted that each unevenly distributed meaning space would favour a particular categorization. This is precisely what they found: in the size condition, the words tended to categorize the objects by size, for instance a word for larger shapes and another for smaller shapes, while in the brightness condition, the emerging categories were aligned with the darker and lighter shapes.

The idea that spurred Perfors and Navarro's study was the Bayesian prediction that the outcomes of an iterated learning chain could be explained by the learners' prior biases. This is most clearly shown in Kalish et al. (2007). These authors run chains of participants who had to implicitly learn mathematical functions: each participant was shown a horizontal bar of a certain length on the screen and had to respond to it by producing a vertical bar of some length. After they had produced their vertical length, they were given feedback as to what the response should have been. The horizontal and vertical magnitudes were, in fact, the x and y values of a function. For instance, in the case of the simplest linear function, $y = x$, they learned to respond so that the longer the horizontal bar, the longer the vertical one should be. They initiated a total of eight transmission chains with different mathematical functions including the above-mentioned positive linear function $y = x$; the inverse negative function $y = 1 - x$; a nonlinear function; and a random correspondence of y values to x values. After nine generations, in all but one of the chains, the function had turned into a positive linear $y = x$, and the remaining one into the negative linear function $y = 1 - x$. These results were interpreted in Bayesian terms, with the posterior distribution of linear functions (the final seven positive linear and one negative linear functions) reflecting the prior (the cognitive

Fig. 5 The words produced by a fourth generation participant in one of the chains Kirby et al. (2008) (Exp. 1). Hyphens have been added for clarity's sake

	Black	Blue	Red	
	wi-ne-kuki wi-nu-kuki wi-ke-kuki	ku-n-kuki hu-ne-kuki ku-ne-kuki	po-ne-kuki pu-ne-kuki po-ne-kuki	○ □ △
	wi-ne-kiko wi-ne-ki wi-kiko	ku-n-kiko hu-ne-kiko ku-ne-kiko	po- -kiko pu- -niko po- -kiko	○ □ △
	wi-ne-ko wu-ne-ko wi-ke-ko	ku-ne-ko hu-ne-ko ku-ne-ko	po-ne-ko pu-ne-ko po-ne-ko	○ □ △

preference for linear functions, especially the positive linear one, Kalish et al. 2004).

Perfors and Navarro’s posterior (final) distribution of languages could not simply be a reflection of (prior) cognitive biases, since this prior was presumably the same for the participants in the size and the brightness conditions. With their ingenious experiment, they showed that cognitive priors are not fixed. Instead, they can be affected by external factors, in this case by the structure of the world, which suggested alternative partitions or categorizations of the meaning space. This experiment is a transparent demonstration of a cultural system—a set of words that are transmitted generation after generation—adaptively responding to the structure of its environment, the meanings that the words refer to.

Iterated learning experiments are designed to establish the role of repeated transmission on language structure and to highlight how cultural information adapts to its own transmission (Kirby et al. 2008). They assume a tight learning bottleneck, with very little training and a pressure to generalize the language learned to express novel, unseen meanings. This may be a valid assumption for language learning, but in many other realms of culture, learning involves extensive teaching and feedback (think of the years of formal education or learning to play an instrument). The differential effect of these two types of learning—without and with feedback—is tested in a study by Tamariz and Kirby (in press). This is, again, an iterated learning experiment, where participants had to look at a drawing for ten seconds and then reproduce it as accurately as they could. The drawing produced by one participant would be the original for the next one, and the initial drawing in all chains was a meaningless doodle (Generation 0 in Fig. 5). In half of the transmission chains, the original drawings were removed from view after the ten seconds (modelling unsupervised, limited training), but in the other half, they remained in full view while the participants copied them (feedback). In this way, the memory element of the transmission could be explored.

In the memory condition, the drawings, as in the previously described whistle and drawing studies, became simpler over the generations (Fig. 5, top). They did so by turning smaller and more streamlined, but also by transforming into conventional

numbers of letters—memorizing “capital R” is much more economical than memorizing the description of a complicated doodle. The drawings in the copy-from-view condition, however, although changing, retained the initial level of complexity and remained meaningless (Fig. 5, bottom). This result provides a clear indication that simplification is caused by keeping information in memory—even if only for a few seconds as in this experiment. Conversely, it shows how different aspects of transmission (or inheritance), such as learning, keeping in memory and reproducing a pattern, affects the structure of the pattern in distinct ways.

2.5 *Conclusions from Experiments and Simulations*

This sample of empirical studies has shown a variety of evolutionary processes in action: inheritance of information as it is passed on from interlocutor to interlocutor and from experienced user to learner; and selection of information patterns that are best adapted to environmental factors such as other patterns, people’s biases and the structure of the world.

We have learned that social communicative interactions are required for the emergence and spread of communicative systems and conventions. We have also seen that when many patterns evolve together, they are influenced by each other and become a system. Finally, language has peculiarities that make it special in several respects: linguistic forms compete for meanings, and they need to be flexible enough to express endless novel meanings during usage; one efficient adaptive solution to this is compositional structure.

The next section integrates this knowledge into a theoretical framework of language evolution.

3 **Elements for a Theory of the Cultural Evolution of Language**

The first part of this chapter reviewed a selection of recent experiments and computer simulations focusing on how they implement cultural-evolutionary processes such as inheritance, variation and selection. This second part outlines a theoretical framework for cultural evolution, centred on communicative systems. The main elements discussed in this framework are, first, *inheritance* of linguistic information: What are the mechanisms of language transmission? Can we talk of replication of linguistic patterns? And second, *selection*: What is the environment where linguistic patterns evolve? What effects do environmental factors have on the patterns?

I will start off by highlighting some high-level commonalities between some of the processes in language evolution and in the origin of life. The beginning of life, before DNA and other complex molecules existed, was characterized by cyclical

chemical reactions involving autonomous replication, or “continued growth and division which is reliant on input of small molecules and energy only” (Szostak et al. 2001). Replication occurred whenever new similar molecules were produced. Variation was brought about by random changes in the molecular structure and by recombination of different molecular parts by horizontal transfer. The feature of this early life system that is relevant to the present discussion is that it did not include translation: the molecules did not code for anything in the way genes today code for proteins. The only “function” of these molecules was self-replication, and the dynamics of the system selected the best replicators: molecular structures that replicated more faithfully increased in frequency and therefore produced even more (faithful) copies of themselves. In the long run, the best replicators would come to prevail. This contrasts with present-day genes, stretches of DNA, which, by virtue of coding for proteins, have functions contributing to the success of the organism that carries them.

The transition between “selection for replicability” only (or the evolution of replicators) and the addition of “selection for function” (or natural selection) came about when the replicating molecules began to code for proteins which, in turn, altered the environment where the molecules replicated. This transition is the third of the major evolutionary transitions proposed by Maynard Smith and Szathmáry (1995). For Woese (1998), it is *the* major transition of life; he calls it “the Darwinian threshold” because it marks the beginning of genes defined by their functions, which constitute the units of natural selection. After the Darwinian threshold, vertical transfer of genetic information leads to an increasingly permanent organismal phylogenetic trace (Woese 1998).

An analogy of these processes in the origin and evolution of language would be the view that humans began to produce vocalizations that carried no symbolic or referential meaning, perhaps similar to birdsong. The “musical protolanguage” hypothesis of language origin (e.g. Darwin 1871; Okanoya 2002; Fitch 2010) does just this. Versions of this hypothesis share the assumption that our hominin ancestors evolved the capacity for *vocal learning*, that is, for faithfully imitating vocal patterns—or, more widely, motor patterns including rhythmical, gestural or vocal sequences. Among our closest relatives, we are the only species capable of (and indeed prone to) imitating behaviour even if it has no apparent function.²

The stage of the origin of life that Woese (1998) would call pre-Darwinian corresponds in the musical protolanguage hypothesis to sounds that are transmitted socially, but which have no communicative function—maybe tunes or dance patterns, hence the “musical” name of this hypothesis. The sounds that were faithfully copied would persist over time, the rest would not: this is selection for replicability. The Darwinian threshold would be crossed when sounds and their combinations began to be produced and understood as meaningful. Now, as well as sounds being selected for replication, certain sound combinations would be

² Vocal learning has been attested in several other distant branches of the tree of life including some species of cetaceans, elephants, bats, song birds, parrots and hummingbirds (Jarvis 2006).

selected for because they conveyed useful meanings, or because they conveyed them well. This is selection for function.

However, the two levels of selection, for faithful replicators and for function do not need to have occurred sequentially either in life or in language (as is proposed to have occurred in the musical protolanguage hypothesis). In fact, a co-evolutionary scenario involving the human capacity for imitative vocal (or more generally, motor) learning, the presence of increasingly complex vocalizations in the environment and co-opting those vocalizations for communicative purposes, in the style of Dor and Jablonka (2010) is equally if not more plausible and has the advantage that it does not require to posit a non-communicative function for the early vocalizations.

In the following sections, I argue for a theoretical approach to language evolution that involves these two levels of evolution in language: on one hand, our species has evolved imitative skills that ensure the faithful replication of the stuff of language—mainly, sounds. On the other hand, combinations of those sounds have functions: we use them to communicate meanings to each other. Communication involves many interrelated factors such as the concepts we entertain and their structure, the alignment of concepts in interlocutors, patterns of social interaction, which may pose selection pressures on the evolution of linguistic items.

3.1 What Evolves in Language Evolution?

The opening paragraphs in this chapter characterized language as behaviour, and what follows is based on this view: I will talk about linguistic patterns, and by that I mean patterns we produce: speech sounds, words, constructions and structural patterns. Linguistic patterns, therefore, do not include the functions of words, constructions, etc. The functions of linguistic items, or “meanings” in an extended sense, include semantic meaning but also all the nuances a word produced in context may convey—the identity and status of the speaker, the degree of formality or informality of the context, the nuances of meaning perceived in the particular context, etc. Those factors, in the current approach, together with speakers and their intentions, constitute an important part of the *environment* where linguistic behavioural patterns evolve. This contrast with Croft’s (2000) model of the evolution of language, where linguistic replicators or “linguemes”, include not only the sounds uttered by speakers but also the meanings of those utterances. Linguemes are linguistic conventions (sound patterns plus their shared meaning) that are replicated each time they are used and are passed on to new generations through usage and learning. The current approach is closer to that of Ritt’s (2004), who concedes that the replication of meaning together with the form is highly problematic and gives a nuanced definition for replicators from which meaning has all but disappeared, leaving only the sounds uttered (or, more specifically, the neural activation patterns that lead to the sounds being uttered). To reiterate, in the present model, linguistic replicators are, more in accordance with Ritt’s proposal, purely behavioural, while meanings are part of the environment where they evolve.

The main argument for putting behaviour at the centre of this theoretical framework is that meanings, the mental representations corresponding to linguistic forms, are much more variable between speakers than the linguistic behavioural form. The number of possible combinations of percepts and concepts must be vastly greater than the number of linguistic patterns found in any one language, if only because the same patterns are reused in multiple and diverse occasions and contexts. Naturally, there is something in common among all the occasions and contexts where the same form is used, but there is also much conceptual information in the brain that is under- or non-specified in linguistic forms. If experience incrementally contributes to the function of each linguistic construction, and if we assume that individual experiential histories are unique, then the individual differences in meaning for each linguistic pattern will be orders of magnitude greater than differences in the corresponding linguistic pattern. We may say that, for each word, there are as many meanings as there are speakers in the language.

The set of all meanings in an individual yields an overall meaning space constituting “a complicated network of similarities overlapping and criss-crossing” (Wittgenstein 1953: 66e). These complex entities cannot be the same for two speakers and may even be different for the same speaker on different occasions. Behavioural linguistic patterns, on the other hand, have similar linguistically relevant features between speakers. It is true that even leaving aside non-linguistic differences such as voice timbre, quality or volume, there is still variation in the forms of words or sounds within the same language. (But “correct speech perception irrespective of the acoustic variation between the different speakers and word context” can be explained by “the existence of such neuronal populations in the human brain that can encode acoustic invariances specific to each speech sound” (Näätänen 2001:1.)

In his model of cultural evolution, Sperber (1996) defends that meanings (mental representations) “are more basic than public ones” (ibid: 78). And yet, he maintains that the cultural transmission of mental information is fundamentally transformational and those transformations are not inheritable. Any stability in cultural (mental) representations across individuals and over time is explained by “attractors” or “points or regions in the space of possibilities, towards which transformations tend to go”.³

Further arguments and evidence for the higher stability and fidelity of transmission of public productions can be found, paradoxically, in the midst of expositions about the primacy of mental culture. In a critique to Sperber (2000), Dennett (2006) claims that public cultural items such as recipes, wheels or renditions of a musical piece can be faithfully transmitted “thanks to the shared norms for (...) analog processes already inculcated in the apprentice”. In Dennett’s argument, however, fidelity relies on the apprentice already being enculturated. Tomasello, Kruger and Ratner (1993) look precisely into enculturation and assert that “human beings ‘transmit’ ontogenetically acquired behaviour and information, both within and across generations, with a much higher degree of fidelity than other animal species”. Richerson

³ An epidemiology of representations: A talk with Dan Sperber. *Edge*, 164, 27 June 2005. <http://www.edge.org/documents/archive/edge164.html#sperber>. Accessed 6 June 2011.

and Boyd, while arguing for mentalistic culture, observe that: “[I]nformation in one person’s brain generates some behaviour—some words, the act of tying a knot, or the knot itself—that gives rise to information in a second person’s brain that generates a similar behaviour. If we could look inside people’s heads, we might find out that *different individuals have different mental representations* of a bowline, even when *they tie it exactly the same way*” (2005: 63–64; my italics). On the same vein, Hodgson and Knudsen note, in their evolutionary approach to economics, that “with habits, replicative similarity is necessarily present at the behavioural level, but unlikely at the neural or genotypic level” (2004: 288). Shennan points out that “the resemblance between the inputs and the public outputs is often very striking” (2002: 47), as illustrated by the continuity observed in many prehistoric pottery traditions (ibid: 47). Another remarkable example of this continuity is the persistence of the same designs and manufacture processes in the Oldowan and Acheulean stone tools in the archaeological record for *over one million years* with negligible modification (although no assumptions can be made about the (lack of) stability of the mental representations of the producers of those tools, since they were hardly human). Not quite as long but equally impressive are the timescales of linguistic items proposed by Pagel (2009), who argues that the origin of the oldest words may be traced back over tens of thousands of years. In the same chapter, he proposes that words, phonemes or syntax constitute “discrete heritable units” (Pagel 2009: 406, Table 1), which are transmitted through “teaching, learning and imitation” (ibid). All these arguments together point to the cultural inheritance of public behaviours and indeed characterize this inheritance as *replication*.

Public cultural manifestations are caused by mental activity at the individual level. The use of “activity”, as opposed to “representations”, is not accidental. Mental representations (Sperber 1996) and Cognitive Causal Chains (Sperber 2006) involve semantic content and relationships. Behaviours and artefacts are the product of the implementation of neural motor instructions, which are in turn caused by other neural activity, perceptual and associative in nature. At this level of analysis, the causality pathways can in principle be established (for instance, with priming experiments that can reveal associations or with methods that provide windows into brain activity). Mental cultural representations, in contrast, are *emergent* from public manifestations. This is notably the case at the individual level, when the patterns of brain connectivity change in response to experience over life-long learning. Like all emergent or complex phenomena, mental representations are sensitive to local conditions and therefore unique and unpredictable—in other words, non-replicable.

3.2 Selection for Replicability in Language

Linguistic (behavioural) patterns can persist over long periods of time, therefore, because they are reproduced faithfully generations after generation. In most of the models reviewed in the first part of this chapter, reproduction is assumed

(e.g. a skill given to the agents in a simulation) or expected (e.g. when human participants are expected to learn and reproduce a typed string of letters). A couple of the experiments, however, acknowledge that learning to produce the signals themselves is not trivial. When the participants in Galantucci (2005) and Verhoef et al. (2011, 2012) were confronted with their props (the distorting writing pad and the slide whistle, respectively), they had to learn the relationship between their movements and the output signals. During the course of the experiments, they became increasingly proficient with their props, and consequently more in control of the structure of the drawings and whistles they produced.

Similarly, human infants need to learn how to control their vocalizations (or signed gestures). Early in their development, they construct a perceptual-motor machinery that allows them to faithfully reproduce specifically the sounds of their language. This machinery develops during babbling. From 5 to 7 months of age, infants tune their motor-articulatory and auditory-perceptual capacities to accurately match the patterns (phonemic categories, intonation patterns) of their ambient language (Braine 1994; Vihman et al. 2009), at the same time as they imitate other motor skills (Iverson et al. 2007; Thelen 1981, cited in Vihman et al. 2009). One proposed mechanism underlying faithful imitation of sounds is an “articulatory filter” (Vihman 1993) whereby sound patterns that the child has already produced during babbling become more perceptually salient. This allows infants to notice frequent patterns in the input speech stream, which prompts further repetition (Vihman et al. 2009). Patterns produced in babbling that are not reinforced by the external input are repeated to a lesser extent, resulting in a repertoire of sounds that resembles that of the ambient language. De Boer’s model, where distinct vowel categories emerged out of feedback between agents, models some aspects of the learning dynamics that goes on during babbling.

As we have seen, the replication of phonemic categories in new speakers can be modelled as blending inheritance (Wedel 2006). The resulting stable categories⁴ translate into a multimodal distribution of acoustic values in the output data produced by speakers. Maye et al. (2002) elegantly demonstrated how distributions of acoustic values in the input translate into functional categories through statistical learning. When they exposed 6- and 8-month-old infants to sounds from a phonetic continuum with a bimodal distribution, the infants were able to discriminate between sounds from both ends of the continuum. Then, the distribution of the sounds was unimodal; however, the infants would not discriminate between the same two sounds. This indicates that the infants inferred two categories from the bimodal distribution and a single category from the unimodal one. It is this sensitivity to the input’s surface statistics that sculpts the fuzzy but distinct, functionally discrete, sound categories. Infants will go on to produce (by blending inheritance) sounds with the same underlying distribution properties as the one they learned, so, over the generations, the statistical properties of phonemic categories are

⁴ They are stable only relatively speaking since, like any replicator in an evolving system, they are subject to mutation.

maintained. Extending this dynamics to a population leads to the emergence of shared systems of phonemes, as modelled in de Boer (2001) and others (Oudeyer 2006; de Boer and Zuidema 2010).

In established languages, selection for replicability may be difficult to detect because an optimal stable state has been reached, but even then it still would act as a stabilizing mechanism, tending to maintain things as they are. In emergent systems, it should lead to the appearance of replicator lineages. An emergent language, Nicaraguan Sign Language, which was spontaneously created by a community of deaf children brought together to a school for the deaf in Managua only a few decades ago, provides a window into the genesis of linguistic systems (e.g. Senghas and Coppola 2001; Senghas et al. 2004; Sandler et al. 2005; Aronoff et al. 2008) and thus gives us the opportunity to examine the forces that operate on the origin of phonemic categories.⁵ In Nicaraguan Sign Language, selection for replicability was at work during the emergence of phonemes and continues to stabilize the existing phonemes. Al-Sayyid Bedouin Sign Language (Aronoff et al. 2008) is another recent sign language, but here, stable phonemic categories are not attested. This may be due to its different population circumstances, though. Nicaraguan Sign Language appeared around 1977 and has now hundreds of Signers. Al-Sayyid Bedouin Sign Language, although it has been around since the 1940s, is used along with spoken language in a smaller mixed community where deaf people are a (sizeable) minority. Perhaps, phonemes will emerge in this language in the future.

During the first year of the life of an infant, the emphasis of language-related learning seems to be directed towards constructing the segmentation of her acoustic-articulatory space that allows her to produce faithful copies of the phonemic categories of the ambient language. As was the case with early life self-replicating molecules, the main “function” of these phonemic categories is self-replication, and the dynamics of the system selects, over generations, for the easiest-to-produce or easiest-to-perceive sounds. Ontogenetically, an infant’s initial babblings have no meaningful content, so individual sounds or their combinations are not selected because of their functions. At this very early stage, Darwinian dynamics with respect to meanings does not exist.

3.3 Selection for Function in Language

The accurate replication of phonemic categories is pivotal for linguistic replication, as the exemplars in these categories are the discrete, replicable and combinatorial units which recombine to form larger linguistic patterns—strings of

⁵ For Kirby (in press), the emergence of duality of pattern, and also of compositionality and the split between content and function words, represents instances of major evolutionary transitions within language.

phonemes, ways to organize strings of phonemes, intonation patterns, etc. These larger patterns are recurrently produced across similar contexts for similar objectives; in other words, they have functions. Functions are defined by the association of a pattern with the contexts where it is produced and the effects that it is perceived to achieve. With each production, the produced pattern—but neither the context nor the function—is replicated. The same pattern can be produced for more or less similar—but rarely identical—functions, which, in turn, will be required in more or less similar—but rarely identical—contexts. Language learning includes the process by which we become able to use the same patterns as fellow speakers for similar functions (in appropriate contexts); but in that process, the only thing that is replicated (or copied with similarity, transfer of information and causality between original and copy) are sounds and sound combinations, nothing else.⁶ Sounds, as we have seen, are replicated thanks to the perceptual and motor learning that takes place early on in an infant's life. (Meaningful) sound combinations are replicated when they are produced in communicative contexts.

The moment sounds or combinations of sounds become symbolically associated with meanings, the system crosses the Darwinian threshold. From that point onwards, as well as sounds being selected for their replicability, sound combinations are selected for because they convey certain meanings, and lineages of words and other constructions can be traced. When linguistic forms compete for meaning niches, we can talk about selection for function.

The mechanism for replication of construction-form replicators involves symbolicity—our capacity to arbitrarily associate a pattern to a meaning (Deacon 1997), which is subsumed in *arbitrary imitation*—our capacity to reproduce of arbitrary symbols observed in others. The main foundation of arbitrary imitation is a kind of imitation variously referred to as “imitation learning” (Tomasello 1996), “true imitation” (Zentall 2006), “observational learning” (Carroll and Bandura 1982), “blind imitation” (Gergely and Csibra 2006) or “complex imitation” (Heyes 2013). In arbitrary imitation, copies of the forms (or means) rather than the functions (or ends) are produced. This is opposed to emulation, where the observer focuses on the goal and employs any means to achieve it, not necessarily the ones used by the model.

Arbitrary imitation is hugely developed in humans, particularly in human children, but not quite there in non-human primates (Tomasello 1996; Whiten et al. 2004). Chimpanzees, for instance, like human children, can imitate complex behaviour sequences in order to achieve a particular goal (Horner et al. 2006). However, if a chimpanzee discovers that an element in the sequence is unnecessary for the goal, or has no function, it tends to stop producing it. In contrast, in the same circumstances, four-year-old children tend to stick to the complete, partially pointless sequence (Horner and Whiten 2005). It is not clear whether

⁶ Of course, in sign languages, it is not sound combinations, but bodily signals that are replicated. And both in spoken and signed languages, some conventional or quasi-conventional co-speech gestures and sounds can also be said to replicate.

children do not analyse the sequence into units, or they do not look for sub-goals, or, even if they do analyse the sequence and realize that an element is unnecessary they nevertheless continue to reproduce it; whatever the exact nature of the process, the result is that the focus of imitation in children are sequences of behaviour. Children's behaviour, therefore, is less rational than that of chimpanzees, but it is more conformist and it implies a high degree of confidence, or trust, that useful information is out there at their disposal. Arbitrary imitation is especially well developed in human children, attested by their tendency to engage in pretend play, as well as by a number of experiments showing that they will imitate intricate actions even when they are obviously over-elaborated for the intended goal (Meltzoff 1988; Horner and Whiten 2005; Lyons et al. 2007; Flynn and Whiten 2008; Whiten et al. 2009). Arbitrary imitation learning is also the mechanism behind the "ratchet effect", which makes cumulative evolution possible in human culture (Tomasello et al. 1993). These studies, in sum, show that humans focus on *means* as they unquestioningly imitate observed arbitrary behavioural patterns, whether they have a utility function or not, other primates focus on *ends* when they only reproduce the actions that (they are persuaded) are functional.

For arbitrary imitation to be possible, two types of abstraction are necessary: abstraction of form from function and abstraction of the signal from the producer of the signal. The first type of abstraction concerns the "arbitrary" part of arbitrary imitation. Arbitrariness is a property of the symbolic associations that link linguistic forms with their meanings (de Saussure [1916] 1983). Forms and meanings are not transparently related to each other, but rather, we simply learn and accept that they are conventionally linked. Apart from language, cultural institutions such as money, democracy or rituals also rely on arbitrary conventions (a bank note or a voting ballot have the value they have because everyone behaves as if they do), while others, like technology, cannot do so.⁷ In order to dissociate a signal—a behaviour—from its function, an arbitrary imitator must be able to *abstract form from function*—or means from ends—that is, decouple an action from its iconic or primary utility function.

The attribution of a novel function to an existing behaviour is not trivial, in fact human adults and children over six find it difficult and display what is called *functional fixedness*: solving a task by using a tool for a novel function is slower if they already had associated the tool with its known utility function (Adamson 1952; Defeyter and German 2003; German and Defeyter 2000). Chimpanzees, incidentally, show extreme functional fixedness (Hanus et al. 2011), while human children under six do not, and are happy to assign new functions to tools that already had a known function. In the experimental game where participants co-opted their movements around the board to communicate the colour of the box where

⁷ The function of a piece of technology is constrained by the properties of its formal structure (German et al. 2007). A hammer is good for hammering and a fishhook is good for catching fish, and they are not interchangeable (and never will be, no matter how much people agree on the opposite).

they would land (Scott-Phillips et al. 2009), functional flexibility happened, not without difficulty, because the players had been given a riddle they had to solve cooperatively and they may have been actively looking for any useful informative cues. The motions around the board were sufficient to meet the limited expressive requirements of the task, namely to distinguish between only four colours. Functional fixedness, therefore, may hinder learning new arbitrary associations. But a cognitive mechanism that may favour arbitrary imitation is pattern completion (Tamariz 2011), which brings about a feeling of relief when an incomplete pattern is completed. This relief is called secondary reinforcement (Miller and Dollard 1941; Osgood 1953) and is noticeable for example while listening to music (Keller and Schoenfeld 1950), when patterns that confirm our expectations bring about pleasure, but patterns that contradict our expectations produce unease. Secondary reinforcement is exacerbated in certain conditions like Tourette syndrome (Prado et al. 2008) or obsessive-compulsive disorder (Rasmussen and Eisen 1992; Summerfeldt 2004). Pattern completion is closely related to the automatization of motor productions, which has also been proposed to have evolved as a facilitator of language production (Deacon 2007). Pattern completion, thus, does away with the necessity of a utility function. A learned pattern is completed for the sake of completing it.

The second type of abstraction required for arbitrary imitation concerns the “imitation” aspect. Arbitrary communication requires that individuals are able to copy behaviour that they have observed in another individual; in other words, they must be able to assume the role of both receiver and producer of behaviour. For this to be possible, they must be able to *abstract the signal from the producer of the signal*. This capacity is called role-reversal imitation (Tomasello 1999) and is much more developed in humans than in other closely related species. One of the most striking examples of a communication system created and learned through interaction by chimpanzees is ontogenetic ritualization (Tomasello and Call 1997). An example of this kind of ritualization is a baby chimp raising its arms and trying to climb on an adult’s back. After this has happened a few times, it is sufficient for the baby to slightly raise its arms for the adult to understand her request and act accordingly. But these learned rituals have limitations: each of the participants has its role and those roles do not change. And the ritual is restricted to this particular pair of individuals who share first-hand experience of the history of the interactions. In ontogenetic ritualization, the behaviour of each of the participants in the interaction is indivisibly attached to its performer. Humans, on the other hand, do role-reversal imitation spontaneously.

It is possible that the two forms of abstraction are manifestations of a single cognitive adaptation to be less rational, overcome logical expectations, accept any sort of incoming information and flexibly integrate patterns in the input even if their cause is not understood—a part of the process of self-domestication proposed by Deacon involving the “de-differentiation of innate predispositions and an increase in the contribution by a learning mechanism” (Deacon 2007: 92).

In the following section, this theoretical framework is supported by evidence from the empirical studies reviewed in the first part of the chapter.

3.4 *Cultural-Evolutionary Dynamics in Language*

The structure of languages is affected by many and varied pressures. Most of them can be categorized as related either to transmission/learning (e.g. cognitive bias relating to production, perception and processing) or to communication/usage (e.g. alignment of concepts in interlocutors, patterns of social interaction or meaning structure). Transmission and communication are intricately intertwined, as the normal way to learn language involves using it communicatively, except in one respect, highlighted above: the sounds of the ambient language are learned thanks to through a unique combination of perceptual-motor and statistical learning in early infancy. The outcome of babbling is the capacity to reproduce the sounds of a language accurately, in other words, a *mechanism of replication* for linguistic sounds. Once this mechanism is in place, the sounds can be used for communication.⁸

Selection for replicability and selection for function may be easy to tell apart in the example of the origin of life, because the replication of “functionless” molecules may be explained by chemical processes, which are distinct from the selective pressures deriving from genetic function. However, in the case of language, it is difficult to find a plausible explanation for the analogous process, namely the repeated imitative production of gestures or vocalizations that do not have a function at the origin of language. But, as pointed out above, selection for replicability and for function do not have to operate (or have evolved) sequentially—they may do so simultaneously in an interactive way. The first experiment described in this chapter, by Scott-Phillips et al. (2009), may be a model of the interaction between the two types of selection in the origin of communication systems. Here, remember, the two players’ characters had to land in rooms of the same colour. At the beginning of a game, when the players were exploring the game, the movements around the rooms were random, and typically they were not copied—although it is not impossible that the fact that one participant started to move gave his partner the idea of moving around too. When they acquired a communicative function, that is, as soon as they were produced and understood as communicative, they stabilized. They began to be faithfully replicated by both players *because* they had a function. In other words, functionality drove replicability.

In many of the experiments described above, the initial state of the system was usually a random set of signals—letter strings, drawings and whistles. This is probably not a good reflection of what the early stages of language were like—unless we accept the musical theory of protolanguage, where we would have a large set or even a system of non-communicative vocal or gestural signals before they took on a communicative role. More likely, signals became communicative

⁸ Notwithstanding the fact that the processes of perfecting the production of the sounds of a language and using vocalizations communicatively usually happen simultaneously for some months or years.

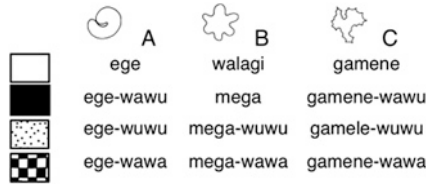


Fig. 6 The sixth and last generation from one of the languages in the Vertical transmission condition in Kirby et al. (in preparation). Hyphens have been added to make the compositional structure more visible. In this language, the first segment refers to the shape (ege means *shape A*; mega *shape B*; and gamene, *shape C*) and the second segment, to the texture (no ending means *white*, wawu means *black*, wuwu means *dotted* and wawa means *checked*). There is one irregular word: walagi, for *shape B, white*; and one irregularity: gamele, instead of gamene, for *shape C*, when it has a *dotted texture*

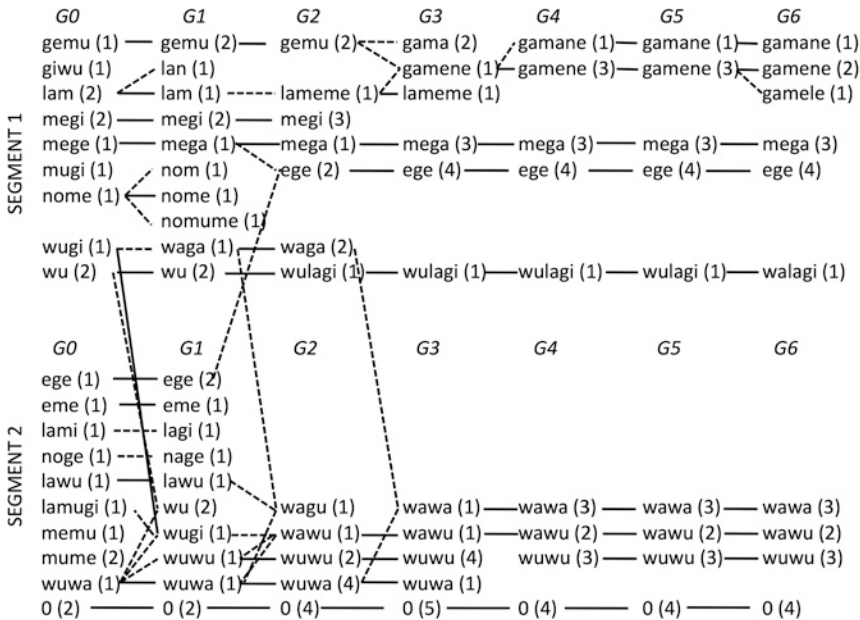


Fig. 7 Coalescent trees of the first and second segments in the words produced at generations G0 to G6 from one of the languages in Kirby et al. (in preparation). The trees were generated following the methods described in Cornish et al. (2009). *Black lines* indicate perfect replication and *dotted lines* indicate recombination or probable descent with modification. The frequency of each segment type at each generation is shown in *brackets*

very early on, as modelled in Scott-Phillips et al. (2009), or in the pictorial games (Garrod et al. 2007; Fay et al. 2010), where the drawings were communicative from the start. Nevertheless, the experimental transmission chains initialized with random signals show how selection for replicability and selection

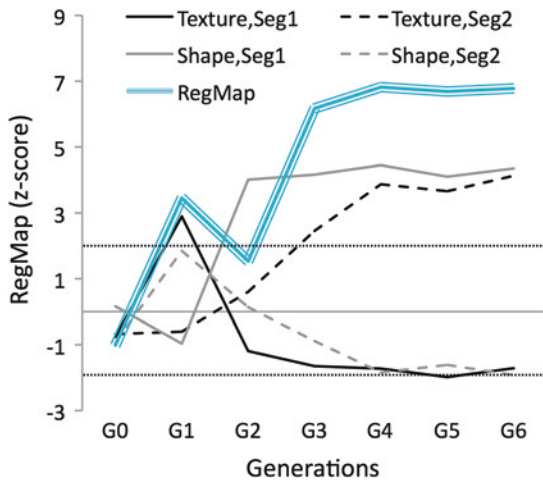


Fig. 8 *RegMap* (double line) and partial *RegMap* values for generations G0 to G6 from the same language in Kirby et al. (in preparation) as in Fig. X. The partial *RegMap* values were calculated following the methods described in Cornish et al. (2009). *RegMap* was calculated by running the same method on the partial *RegMaps*. Z-scores calculated with a Monte Carlo simulation ($N = 5000$). Values near 0 indicate random or irregular mappings. Values above 1.96 indicate the mappings are significantly more regular than expected by chance, while values below -1.96 indicate they are significantly less regular than expected by chance

for function transform randomness into structure. Thus, in languages like those in emerging in Kirby et al. (2008, in preparation), we see the increasing prevalence of more reproducible and increasingly meaningful signal units. The following analysis on one of the artificial language families generated in Kirby et al. (in preparation) clearly illustrates the simultaneous action of selection for replicability and selection for function. Figure 6 shows the final generation of this particular language chain, which began with a random language and was learned used communicatively by six consecutive pairs of players.

The coalescent tree in Fig. 7 shows the evolution of the first and second segments in the language from the initial random language, to the language produced by the sixth generation, and illustrates the emergence of stable replicators. The initial languages contain many different segments, both in first and second position. As the languages evolve, the segments mutate, blend and move from one position to another, while they decrease in number. The surviving segments have higher frequencies (they are reused in several words) and become increasingly stable towards the latter generations, where mutations and position changes are almost non-existent. It is also interesting to note that wild mutations do not occur. The players in the experimental game, even when they could not remember the words for an object, did not invent a totally new word, or introduce new letters; they behaved in a very conformist way (even though they were not required to) and produce only words similar to the set they had been trained on.

The effects of selection for function are clear in Fig. 8, which plots a measure of the level of compositional structure between word segments and meanings (Fig. 6). Partial *RegMap* quantifies the confidence that a segment is consistently associated with a meaning. The graph shows how, from generation G2 onwards, the first segment is clearly, and increasingly stably, associated with shape and the second segment, with texture. The overall *RegMap* value measures the overall confidence that each segment is consistently associated with a meaning in a one-to-one, unambiguous relationship.

In this particular language, which referred to meanings with two features—shape and texture—the signals have split into two meaningful units, with the first one adapted to conveying shape and the second one to conveying texture. In the languages from the experiments in Kirby et al. (2008), where there were three dimensions of meaning (shape, colour and motion), signals split into three meaningful units, each adapted to one dimension (Cornish et al. 2009).

The fitness of a letter-string replicator in these languages (the likelihood that it would be reproduced by the next generation) was determined by how memorable it was, which in turn depended on (a) replication factors, e.g. how easy it was to produce, or how similar it was to letter strings in the native language of the players; and (b) functional factors, e.g. how meaningful it was, or how reliably it was associated with a meaning dimension. The effects of replication factors are apparent in Verhoef et al.'s (2011, 2012) whistle experiments, where the final, evolved whistles were easier to produce than the initial ones. Functional factors are apparent in all the iterated learning of miniature artificial language experiments, with adaptation of forms to meaning space structure being most obvious in Perfors and Navarro's (2011) study where word categories aligned with either the size or the colour of the square objects they had to denote, depending the salience of the differences in square size of colour.

The (functional) fitness of linguistic patterns is also affected by their being associated with certain social values or social identities (Labov 2001; Croft 2000; Richerson and Boyd 2005) or having an iconic relationship with a meaning, as illustrated by the paradigmatic case of words that are cross-culturally preferred to designate a rounded figure (like “bouba” or “maluma”) or a spiky figure (like “kiki” and “takete”) (Kohler 1929; Ramachandran and Hubbard 2001). Some of the fitter drawing systems produced in the distorting-pad to denote the different rooms in Galantucci's (2005) experiments were iconically related to their meanings: lines on either side of the pad referred to rooms on the left or the right of the board. The fitness of a linguistic pattern is also influenced by how often speakers need or wish to refer to their meaning. For instance, the English lexical form “oil-lamp” is not very fit nowadays, as its referent ceased to be frequent in the homes of English speakers. Conversely, the appearance of the Internet has selected for the corresponding word form “Internet”, which is now infinitely more frequent than only two or three decades ago. The form “gay”, on the other hand, used to be selected for through its association with the meaning “happy”, whereas now it is probably even fitter because of its connection with the commonly expressed concept of “homosexual”. Finally, a trait that enhances fitness specifically in

communication systems is being distinct from other linguistic pattern replicators: the drawings produced in the distorting pads in Galantucci (2005) or the words for the object in the communicative games in Kirby et al. (in preparation) were functional and therefore had higher chances of being reused, only if they were distinct from each other. Each drawing, or each word, in a system adapted specifically to one of the meaning niches available.

Natural languages are culturally transmitted institutions and therefore have to be continuously learned by new speakers. Human learners are able to faithfully learn the sounds of their language during babbling and subsequently reproduce them accurately. The most adaptive sounds are those that are easiest to learn and reproduce by the mechanisms involved in babbling. Languages persist because speakers use them for their communicative purposes. The most adaptive linguistic (sound) patterns are those that best convey relevant meanings. The meanings, the speakers and their cognitive capacities and communicative needs are the environment where linguistic replication, innovation and selection take place.

4 Conclusion

This chapter has presented a model of the cultural evolution of language based on mechanisms that are attested in experiments. The main argument for replication stems from the claim that functions help stabilize arbitrary forms. Theories of cultural evolution have not managed to find a consensual framework and I believe this is because they were focusing on the most interesting part of culture: shared social constructs and values, etc. that exist in people's minds. Such mental representations are not faithfully replicated, they do not "leap across brains", and these have constituted serious problems with theories such as memetics. Ideas, values and cultural institutions continue to exist and evolve because the behaviours that give rise to them are faithfully replicated by generation after generation of humans. But because mental representations are emergent from individual experiential paths, they cannot replicate. They may be similar in the same way as the precise paths of two birds in a flock or the noses of grandfather, father and grandson are similar; they belong to the same *kind* of paths and noses, but each is unique. In this chapter, I have described mechanisms for the replication of the public manifestations of linguistic information—and the same mechanisms could well be at work in other cultural institutions, from greetings to money or justice. Culture, including language, exists because of human brains and the knowledge, beliefs and values that emerge in them. But culture, including language, would not exist without human bodies—hands, mouths, arms—that replicate the public behavioural and material expressions of mental constructs.

References

- Adamson RE (1952) Functional fixedness as related to problem solving: a repetition of three experiments. *J Exp Psychol* 44:288–291
- Aronoff M, Meir I, Padden C, Sandler W (2008) Roots of linguistic organization in a new language. In: Arbib M, Bickerton D (eds) Special Issue on Holophrasis, Compositionality and Protolanguage. *Interact Stud* 9(1):131–150
- Beckner C, Blythe R, Bybee J, Christiansen MH, Croft W, Ellis NC, Holland J, Ke J, Larsen-Freeman D, Schoenemann T (2009) Language is a complex adaptive system: position paper. *Lang Learn* 59(S1):1–26
- Bentley RA, Hahn MW, Shennan SJ (2004) Random drift and culture change. *P Roy Soc Lond B Bio* 271(1547):1443–1450
- Blythe R, Croft W (2009) The speech community in evolutionary language dynamics. *Lang Learn* 59(S1):47–63
- Boyd R, Richerson PJ (1995) *Culture and the evolutionary process*. University of Chicago Press, Chicago
- Braine MDS (1994) Is nativism sufficient? *J Child Lang* 21:9–31
- Brighton H, Smith K, Kirby S (2005) Language as an evolutionary system. *Phys Life Rev* 2:177–226
- Brown JE (2012) *The evolution of symbolic communication: an embodied perspective*. Unpublished PhD thesis. The University of Edinburgh
- Carroll WR, Bandura A (1982) The role of visual monitoring in observational learning of action patterns. *J Motor Behav* 14:153–167
- Cavalli-Sforza LL, Feldman MW (1981) *Cultural transmission and evolution: a quantitative approach*. Princeton University Press, Princeton (Monographs in Pop Biol 16)
- Cornish H, Tamariz M, Kirby S (2009) Complex adaptive systems and the origins of adaptive structure: what experiments can tell us. Special issue on language as a complex adaptive system. *Lang Learn* 59(4S1):187–205
- Croft W (2000) *Explaining language change: an evolutionary approach*. Longman, Harlow
- Darwin C (1871) *The descent of man and selection in relation to sex*. John Murray, London
- Dawkins R (1976) *The selfish gene*. Oxford University Press, Oxford
- Deacon TW (1997) *The symbolic species: the co-evolution of language and the brain*. Norton, New York
- Deacon TW (2007) Multilevel selection and language evolution. In: BH Weber BH, Depew DJ (eds) *Evolution and learning: the Baldwin effect reconsidered*. MIT Press, Cambridge, MA
- de Boer B (2001) *The origins of vowel systems*. Oxford University Press, Oxford
- de Boer B (2000) Self-organisation in vowel systems. *J Phon* 28(4):441–465
- de Boer B (1999) Evolution and self-organisation in vowel systems. *Evol Communication* 3(1):79–103
- de Boer B, Zuidema W (2010) Multi-agent simulations of the evolution of combinatorial phonology. *Adapt Behav* 18(2):141–154
- Dennett D (2006) From typo to thinko: when evolution graduated to semantic norms. In: Levinson S, Jaissou P (eds) *Evolution and culture, a Fyssen Foundation symposium*. MIT Press, Cambridge
- de Saussure F ([1916] 1983) *Course in general linguistics*. Duckworth, London
- Defeyter MA, German TP (2003) Acquiring an understanding of design: evidence from children's insight problem solving. *Cognition* 89:133–155
- Dor D, Jablonka E (2010) Plasticity and canalization in the evolution of linguistic communication: An evolutionary developmental approach. In: Larson L, Deprez V, Yamakido H (eds) *The evolution of human language*. Cambridge University Press, Cambridge
- Dunn M, Terrill A, Reesink G, Foley RA, Levinson SC (2005) Structural phylogenetics and the reconstruction of ancient language history. *Science* 309(5743):2072–2075
- Fay N, Garrod S, Roberts L, Swoboda N (2010) The interactive evolution of human communicative systems. *Cognitive Sci* 34:351–386
- Fitch WT (2010) *The evolution of language*. Cambridge University Press, Cambridge

- Flynn E, Whiten A (2008) Cultural transmission of tool use in young children: a diffusion chain study. *Soc Dev* 17(3):699–718
- Galantucci B (2005) An experimental study of the emergence of human communication systems. *Cognitive Sci* 29(5):737–767
- Garrod S, Fay N, Lee J, Oberlander J, MacLeod T (2007) Foundations of representation: where might graphical symbol systems come from? *Cognitive Sci* 31:961–987
- Gell-Mann M (1992) Complexity and complex adaptive systems. In: Hawkins JA, Gell-Mann M (eds) *The evolution of human languages*. Addison-Wesley, Reading
- Gergely G, Csibra G (2006) Sylvia's recipe: the role of imitation and pedagogy in the transmission of human culture. In: Enfield NJ, Levinson SC (eds) *Roots of human sociality: culture, cognition, and human interaction*. Berg Publishers, Oxford
- German TP, Defeyter MA (2000) Immunity to functional fixedness in young children. *Psychon Bull Rev* 7:707–712
- German TP, Truxaw D, Defeyter MA (2007) The role of information about “convention”, “design”, and “goal” in representing artificial kinds. *New Dir Child Adolesc Dev* 115:69–81
- Godfrey-Smith P (2000) The replicator in retrospect. *Biol Philos* 15(3):403–423
- Hanus D, Mendes N, Tennie C, Call J (2011) Comparing the performances of apes (*Gorilla gorilla*, *Pan troglodytes*, *Pongo pygmaeus*) and human children (*Homo sapiens*) in the floating peanut task. *PLoS ONE* 6(6):e19555
- Heyes CM (2013) What can imitation do for cooperation? In: Calcott B, Joyce R, Sterelny K (eds) *Signalling, commitment and emotion*. MIT Press, Harvard
- Herzog HA, Bentley RA, Hahn MW (2004) Random drift and large shifts in popularity of dog breeds. *P Roy Soc Lond B Bio* 271:S353–S356
- Hodgson GM, Knudsen T (2004) The Firm as an interactor: firms as vehicles for habits and routines. *J Evol Econ* 14(3):281–307
- Horner V, Whiten A (2005) Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Anim Cogn* 8:164–181
- Horner V, Whiten A, Flynn E, de Waal FBM (2006) Faithful copying of foraging techniques along cultural transmission chains by chimpanzees and children. *PNAS* 103:13878–13883
- Hull DL (1988) *Science as a process: an evolutionary account of the social and conceptual development of science*. University of Chicago Press, Chicago
- Iverson JM, Hall AJ, Nickel L, Wozniak RH (2007) The relationship between onset of reduplicated babble and laterality biases in infant rhythmic arm movements. *Brain Lang* 101:198–207
- Jarvis ED (2006) Evolution for and against vocal learning in birds and mammals. *Ornithol Sci* 5:5–14
- Kalish ML, Griffiths TL, Lewandowsky S (2007) Iterated learning: intergenerational knowledge transmission reveals inductive biases. *Psychon Bull Rev* 14:288–294
- Kalish M, Lewandowsky S, Kruschke J (2004) Population of linear experts: knowledge partitioning and function learning. *Psychol Rev* 111:1072–1099
- Keller R (1994) *On language change the invisible hand in language*. Routledge, London
- Keller FS, Schoenfeld WN (1950) *Principles of psychology: a systematic text in the science of behavior*. Appleton-Century-Crofts, New York
- Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge
- Kirby S (2001) Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE J Evol Comput* 5:102–110
- Kirby S, Cornish H, Smith K (2008) Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *PNAS* 105(31):10681–10686
- Kirby S, Tamariz M, Cornish H, Smith K, Compression and communication drive the evolution of language (in prep)
- Kohler W (1929) *Gestalt psychology*. Liveright, New York
- Labov W (2001) *Principles of linguistic change, volume II: social factors*. Blackwell, Oxford
- Lyons DE, Young AG, Keil FC (2007) The hidden structure of overimitation. *PNAS* 104:19751–19756

- Maye J, Werker JF, Gerken L (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82(3):B101–B111
- Maynard Smith J, Szathmáry E (1995) *The major transitions in evolution*. Freeman, Oxford
- Whiten A, McGuigan N, Marshall-Pescini S, Hopper LM (2009) Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philos Trans Roy Soc [Biol]* 364(1528):2417–2428
- Meltzoff AN (1988) Infant imitation after a 1-week delay: long-term memory for novel acts and multiple stimuli. *Dev Psych* 24:470–476
- Mesoudi A, Whiten A (2008) The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philos Trans Roy Soc [Biol]* 363:3489–3501
- Miller NE, Dollard J (1941) *Social learning and imitation*. Yale University Press, New Haven
- Nääätänen R (2001) The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38:1–21
- Nettle D (1999) *Language diversity*. Oxford University Press, Oxford
- Odling-Smee FJ, Laland KN, Feldman MW (2003) *Niche construction: the neglected process in evolution*. Princeton University Press, Princeton and Oxford
- Okanoya K (2002) Sexual display as a syntactical vehicle: the evolution of syntax in birdsong and human language through sexual selection In: Wray A (ed) *The transition to language*. Oxford University Press, Oxford
- Osgood CE (1953) *Method and theory in experimental psychology*. Oxford University Press, New York
- Oudeyer PY (2006) *Self-organization in the evolution of speech*. Oxford University Press, Oxford
- Page M (2009) Human language as a culturally transmitted replicator. *Nature Rev Genet* 10:405–415
- Perfors A, Navarro DJ (2011) Language evolution is shaped by the structure of the world: an iterated learning analysis. In: Carlson L, Hoelscher C, Shipley TF (eds) *Proceedings of the 33rd annual conference of the cognitive science society*. Cognitive Science Society, Austin
- Pierrehumbert JB (2012) The dynamic lexicon In: Cohn A, Huffman M, Fougeron C (eds) *Handbook of Laboratory Phonology*. Oxford U Press
- Pinker S, Bloom P (1990) Natural-language and natural-selection. *Behav Brain Sci* 13(4):707–726
- Povinelli D, Bering J, Giambone S (2003) Chimpanzee ‘pointing’: another error of the argument by analogy? In: Kita S (ed) *Pointing: where language, culture, and cognition meet*. Erlbaum, Hillsdale
- Prado HS, Rosario MC, Lee J, Hounie AG, Shavitt RG, Miguel EC (2008) Sensory phenomena in obsessive-compulsive disorder and tic disorders: a review of the literature. *CNS Spectr* 13(5):425–432
- Ramachandran V, Hubbard E (2001) Synaesthesia: a window into perception, thought and language. *J Consciousness Stud* 8(1):3–34
- Rasmussen SA, Eisen JL (1992) The epidemiology and differential diagnosis of obsessive compulsive disorder. *J Clin Psychiat* 53:4–10
- Richerson PJ, Boyd R (2005) *Not by genes alone: how culture transformed human evolution*. University of Chicago press, Chicago
- Ritt N (2004) *Selfish sounds: a darwinian approach to language change*. Cambridge University Press, Cambridge
- Sandler W, Meir I, Padden C, Aronoff M (2005) The emergence of grammar: systematic structure in a new language. *PNAS* 102(7):2661–2665
- Scott-Phillips TC, Kirby S (2010) Language evolution in the laboratory. *Trends Cogn Sci* 14(9):411–417
- Scott-Phillips TC, Kirby S, Ritchie GRS (2009) Signalling signalhood and the emergence of communication. *Cognition* 113(2):226–233
- Senghas A, Coppola M (2001) Children creating language: how Nicaraguan Sign Language acquired a spatial grammar. *Psych Sci* 12(4):323–328
- Senghas A, Kita S, Özyürek A (2004) Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science* 305(5691):1779–1782

- Shennan S (2002) *Genes, memes and human history: darwinian archaeology and cultural evolution*. Thames and Hudson, London
- Sperber D (1996) *Explaining culture: a naturalistic approach*. Blackwell, Oxford
- Sperber D (2000) An objection to the memetic approach to culture In Aunger R (ed) *Darwinizing culture: the status of memetics as a science*. Oxford University Press, Oxford
- Sperber D (2006) Conceptual tools for a naturalistic approach to cultural evolution In: Levinson SC, Jaisson P (eds) *Evolution and culture: a Fyssen Foundation symposium*. MIT Press, Cambridge, pp 147–165
- Steels L (1996) Emergent adaptive lexicons. In: Maes P (ed) *From animals to animals 4, proceedings of the fourth international conference on simulation of adaptive behavior*. MIT Press, Cambridge
- Steels L (1998) The origins of syntax in visually grounded robotic agents. *Artif Intell* 103:1–24
- Steels L (2003) Evolving grounded communication for robots. *Trends Cogn Sci* 7(7):308–312
- Steels L (2006) Experiments on the emergence of human communication. *Trends Cogn Sci* 10(8):347–349
- Summerfeldt LJ (2004) Understanding and treating incompleteness in obsessive compulsive disorder. *J Clin Psychol* 60:1155–1168
- Szostak J, Bartel D, Luisi P (2001) Synthesizing life. *Nature* 409:383–390
- Tamariz M (2011) Could arbitrary imitation and pattern completion have bootstrapped human linguistic communication? *Interact. Stud.* 12(1):36–62
- Tamariz M, Kirby S, Culture: copying, compression and conventionality. *Cognitive Sci* (in press)
- Tamariz M, Ellison ME, Barr D, Fay N, Evidence for selection in the cultural evolution of human communication systems (in prep)
- Thelen E (1981) Rhythmical behavior in infancy: an ethological perspective. *Dev Psychol* 17(3):237–257
- Tomasello M (1996) Do apes ape? In: Heyes CM, Galef BG Jr (eds) *Social learning in animals: the roots of culture*. Academic Press, London
- Tomasello M (1999) *The cultural origins of human cognition*. Harvard University Press, Cambridge
- Tomasello M, Call J (1997) *Primate cognition*. Oxford University Press, Oxford
- Tomasello M, Kruger A, Ratner H (1993) Cultural learning. *Behav Brain Sci* 16:495–552
- Trudgill P (2004) *New-dialect formation: the inevitability of colonial Englishes*. Edinburgh University Press, Edinburgh
- Verhoef T, Kirby S, Padden C (2011) Cultural emergence of combinatorial structure in an artificial whistled language. In: Carlson L, Holscher C, Shipley T (eds) *Proceedings of the 33rd annual conference of the cognitive science society*. Cognitive Science, Austin, pp 483–488
- Verhoef T, de Boer B, Kirby S (2012) Holistic or synthetic protolanguage: evidence from iterated learning of whistled signals. In: Scott-Phillips T, Tamariz M, Cartmill EA, Hurford JR (eds) *The evolution of language: proceedings of the 9th international conference*. World Scientific, Hackensack
- Vihman MM, DePaolis RA, Keren-Portnoy T (2009) Babbling and word: a dynamic systems perspective on phonological development. In: Bavin EL (ed) *The cambridge handbook of child language*. Cambridge University Press, Cambridge
- Vihman MM (1993) Variable paths to early word production. *J Phonet* 21:61–82
- Wedel A (2006) Exemplar models, evolution and language change. *Linguist Rev* 23:247–274
- Whiten A, Horner V, Litchfield C, Marshall-Pescini S (2004) How do apes ape? *Learn Behav* 32:36–52
- Wittgenstein L (1953) *Philosophical investigations*. Blackwell, Oxford
- Woese CR (1998) The universal ancestor. *PNAS* 95:6854–6859
- Zentall TR (2006) Imitation: definitions, evidence and mechanisms. *Anim Cogn* 9:1435–1448