Language Adapts: Exploring the Cultural Dynamics of Iterated Learning

Hannah Cornish, M.A.(Hons), MSc.

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy

to

Linguistics and English Language

Philosophy, Psychology and Language Sciences

The University of Edinburgh

2011

Abstract

Human languages are not just tools for transmitting cultural ideas, they are themselves culturally transmitted. This single observation has major implications for our understanding of how and why languages around the world are structured the way they are, and also for how scientists should be studying them. Accounting for the origins of what turns out to be such a uniquely human ability is, and should be, a priority for anyone interested in what makes us different from every other life-form on Earth.

The way the scientific community thinks about language has seen considerable changes over the years. In particular, we have witnessed movements away from a purely descriptive science of language, towards a more explanatory framework that is willing to embrace the difficult questions of not just how individual languages are currently structured and used, but also how and why they got to be that way in the first place. Seeing languages as historical entities is, of course, nothing new in linguistics. Seeing languages as complex adaptive systems, undergoing processes of evolution at multiple levels of interaction however, is.

Broadly speaking, this thesis explores some of the implications that this perspective on language has, and argues that in addition to furthering our understanding of the processes of biological evolution and the mechanisms of individual learning required specifically for language, we also need to be mindful of the less well-understood cultural processes that mediate between the two. Human communication systems are not just direct expressions of our genes. Neither are

they independently acquired by learners anew at every generation. Instead, languages are transmitted culturally from one generation to another, creating an opportunity for a different kind of evolutionary channel to exist. It is a central aim of this thesis to explore some of the adaptive dynamics that such a cultural channel has, and investigate the extent to which certain structural and statistical properties of language can be directly explained as adaptations to the transmission process and the learning biases of speakers.

In order to address this aim, this thesis takes an experimental approach. Building on a rich set of empirical results from various computational simulations and mathematical models, it presents a novel methodological framework for exploring one type of cultural transmission mechanism, iterated learning, in the laboratory using human participants. In these experiments, we observe the evolution of artificial languages as they are acquired and then transmitted to new learners. Although there is no communication involved in these studies, and participants are unaware that their learning efforts are being propagated to future learners, we find that many functional features of language emerge naturally from the different constraints imposed upon them during transmission.

These constraints can take a variety of forms, both internal and external to the learner. Taken collectively, the data presented here suggest several points: (i) that iterated language learning experiments can provide us with new insights about the emergence and evolution of language; (ii) that language-like structure can emerge as a result of cultural transmission alone; and (iii) that whilst structure in these systems has the appearance of design, and is in some sense 'created' by intentional beings, its emergence is in fact wholly the result of non-intentional processes. Put simply, cultural evolution plays a vital role in language. This work extends our framework for understanding it, and offers a new method for investigating it.

Declaration

I declare that this thesis was composed by myself, and that the work contained herein is my own except where explicitly stated otherwise in the text. This work has not been submitted for any other degree or professional qualification except as specified.

Hannah Cornish

Edinburgh, September 2010

Acknowledgements

I have learned two very significant things over the course of completing this doctorate:

1. There is always something more exciting to do than sit down to write your thesis. For that I thank my wonderful friends and family (you know who you are), without whom I would have gone mad years ago.
2. There is never anything more important to do than sit down to write your thesis. For that I thank my amazing supervisors, Simon Kirby and Mónica Tamariz, without whom I would also have gone mad years ago.

This project, like many projects, was begun by a good idea: why don't we turn a simulation into an experiment? As much as I would be honoured to have thought up this good idea by myself, there is something quite satisfying about the fact that I did not. In fact, no one individual did. Much like the languages that emerge over time in my experiments, this particular idea seemed to evolve of its own accord.

I can remember the moment exactly. There were three agents involved - myself, Simon Kirby and Kenny Smith - and there was a number of constraints that needed to be adapted to. I needed to find a project for my MSc that fulfilled the rules and regulations set out by the board of examiners. I wanted to do a computational model of language evolution. The rules said I had to do an experiment. "No, no, no!" I said. I *wanted* to study language evolution! But I *had* to do an experiment.

Cue awkward silence. Commence scratching heads. Pause to let disappointment drain away. Language evolution. Experiment. Want. Need.

Oh.

And there it was. The beginning of a crazy adventure that would lead me to conferences in Rome, Barcelona and Utrecht, to a geek retreat in the hills of Santa Fe, and to a film studio in London. To me presenting some of these ideas to Fellows of the Royal Society, members of the Royal family, and five-year old children all on the same day. To me spending hours upon hours in basement laboratories all over campus, meeting my datapoints face to face. To me having the best and worst time of my life to date.

So now, it is time to own up and acknowledge those who have helped make everything thus. It is most definitely true that this work would not have been possible without the efforts of many kind-hearted souls. The faults are still all mine, but so many parts of this thesis have been knowingly or unknowingly improved, inspired, cajoled or helped along by clever ideas or practical assistance from folk that I feel I simply must mention them by name: Peter Bell, Mike Bennett, Morag Brown, Lynsey Buchanan, Christine Caldwell, Morten Christiansen, Rob Clark, Chrissy Cuskley, Dan Dediu, Eddie Dubourg, Adam Duguid, the Economic and Social Science Research Council, Vanessa Ferdinand, Francesca Filiaci, Molly Flaherty, Bruno Galantucci, Ellen Gurman-Bard, Stefan Hoefler, Jim Hurford, Joolie Kasmire, Katie Keltie, Simon Kirby, Cyprian Laskowski, Pippa Line, Karen Ludke, Cristina Matthews, Cassie Mayo, Manuel Marques-Pita, Keelin Murray, Toni Noble, Mits Ota, Justin Quillinan, Gareth Roberts, Sean Roberts, Thom Scott-Phillips, Barbora Skarabela, Andrew Smith, Kenny Smith, Mónica Tamariz, Carrie Theisen, Rob Truswell, Dan Wedgwood, Sharon Whyte, and James Winters. Many thanks, beers and love to you all.

Finally, to my Mum and Dad: for giving all they could, more than they should, and for always letting me be me. You are both amazing, and I hope that you know that.

For Betty and Young Ed. Wombats together, and missed.
x.x.x

Contents

List of Tables

List of Figures

Chapter One

Introduction

> [I]t cannot fail to occur to us as an interesting question, by what gradual steps the transition has been made from the first simple efforts of uncultivated nature, to a state of things so wonderfully artificial and complicated. Whence has arisen that systematical beauty which we admire in the structure of a cultivated language; that analogy which runs through the mixture of languages spoken by the most remote and unconnected nations; and those peculiarities by which they are all distinguished from each other?
>
> — Dugald Stewart (1858)

When we see structure in our surroundings, it is only natural to question the origins of that structure. In spite of its early date, the quote by Dugald Stewart (1753-1828) that begins this chapter anticipates many of the challenges still faced by modern linguists today. How do we explain the 'systematical beauty' that we see in the structure of language? How are we to reconcile both the similarities ('analogy') and the surprising amount of variation ('peculiarities') exhibited amongst the languages of the world? As one would expect, there have been many attempts at addressing Stewart's questions[1] over the intervening years, yet they still remain as open to debate today as they were 150 years ago. This thesis focuses only on the question of the origins of structure in language. In short: why is language structured the way it

[1] The sentiments echoed in this quote are not just those of Dugald Stewart himself, but were also shared by his contemporary, the great economist Adam Smith (1723-1790), whose memoirs Stewart was collecting.

is, and not some other way? It turns out that the answer to this question might also shed light on the other features of language that puzzle us.

1.1 Background

Although communication systems are abundant in nature, one of the things that makes humans different from other animals is that we use language for more than just communication (Dennett, 1995; Jackendoff, 1996; Tomasello, 1999; Lupyan *et al.*, 2007). Not only this, but language itself has some unique properties not found in other systems. In particular, human language is both open-ended (allowing infinite expression of an unlimited set of concepts) and highly variable (Hurford *et al.*, 1998; Evans & Levinson, 2009; Fitch, 2010). If language is the trait that separates us from other animals, then in order to understand what makes us human, we need to understand language (Christiansen & Kirby, 2003). However, in order to fully comprehend a complex phenomenon like language we need to understand it from many different perspectives. It is not enough to simply know how it works, we also need to understand how it came to be that way. In other words, we need to understand how it evolved.

This is not a trivial exercise. Language is both a behavioural skill rooted in human biology and a cultural entity. This means that when it comes to the topic of 'language evolution', we could be referring to: (i) the evolution of the mechanisms responsible for language, or (ii) the evolution of languages themselves. In actual fact, we are concerned with both, all of the time. Although this thesis primarily concerns itself with trying to explain language evolution in the sense implied in (ii), I hope to argue that insights gained from this area can actually help us to identify where we should be focusing our attention on in terms of explaining (i).

The actual study of human language has had a long history. As a result, the way that the scientific community thinks about language has undergone considerable changes over the years. Explanations attributing language to the work of some divine creator have made way for accounts focusing on understanding language in functional, cognitive and behavioural terms. Of particular significance to the work presented here is the increasing movement within mainstream linguistics away from a purely descriptive science of language, towards a more explanatory framework that is willing to embrace the difficult questions of not just how individual languages are currently structured and used, but also how and why they got to be that way in the first place. From being a niche field a decade or two ago, evolutionary linguistics is now booming (Zuidema, 2005).

Part of the reason for this is because it is still easy to appreciate Dugald Stewart's fascination with the mystery of language. The puzzle he sets out has not received a completely satisfying explanation since most theories tend to view language as a phenomenon entirely encapsulated within the individual speaker-hearer (e.g. Chomsky, 1975). Explaining how language evolved in this view amounts to explaining how the brain mechanisms that support language evolved (e.g. Pinker & Bloom, 1990). This underplays the important role of cultural and social interaction between populations of speaker-hearers. One important observation to be made is that languages are not just tools for transmitting cultural ideas, they are themselves culturally transmitted (Brighton *et al.,* 2005). This fact has interesting implications; namely that the process of cultural transmission has an explanatory role to play in the emergence of key structural features of language (Anderson, 1973; Hurford, 1990; Kirby, 1999).

1.2 Thesis Aims

This thesis takes seriously the idea that processes of cultural transmission can explain the emergence of some (perhaps all) of the key structural properties, such as compositionality, duality of patterning, systematicity and possibly recursion, that underlie the features of language like open-endedness and variability mentioned earlier (see §2.1). One of the major contributions that it makes to the field lies in offering up a new experimental methodology to test claims made about the cultural transmission of language. To date, most of what we know about the cultural evolution of language has come from mathematical models and computer simulations (e.g. Hare & Elman, 1995; Batali, 1998; Kirby & Hurford, 2002; Steels, 2003; Brighton *et al.,* 2005; Griffiths & Kalish, 2007; Kirby et al., 2007). What makes the work presented here unique is that it involves obtaining empirical results from populations of human learners, via an experimental paradigm known as human iterated language learning. Here, artificial languages are transmitted between learners under controlled laboratory conditions, allowing researchers to track the changes that take place over time.

One of the central themes which runs throughout what follows is that it is time to start studying language evolution in the laboratory like this. However, this is not to say that there is no room for existing methods of investigation. On the contrary, if we want to make progress in understanding a phenomenon as complex as this, we actually need a greater degree of synthesis and communication between practitioners of different empirical approaches. One factor which is already apparent in the literature is that there seems to be a divide between computational/ mathematical models and other kinds of empirical research. This division does not solely exist within the relatively small field of language evolution. The same situation exists amongst researchers interested in cultural evolution more generally. In this quote from Barrett *et al.* (2002), we could easily replace the word 'culture' with 'language' to make this point:

> "The last few decades have seen the development of two quite independent paradigms in the evolutionarily-informed study of culture. One of these has focused on building mathematical models of the process of cultural transmission (in effect studying inheritance mechanisms); the other has had a more empirical focus, being principally concerned with the adaptiveness of culture." (p351-352)

Barrett et al. (2002) suggest that the reason for this split is in part due to the fact that both strands pose very different research questions. However, they also attribute a degree of blame for the lack of integration to the fact that computational and mathematical modelling still remains mysterious and poorly understood outside of those practising it. Nevertheless, much can be gained by bringing these two different approaches closer together. The experimental framework presented in this thesis represents an explicit attempt to do just this. To describe them as experiments inspired by computational models of iterated learning is an understatement: they are more like actual simulations of iterated learning instantiated in humans, rather than artificial agents.

This makes the experiments themselves somewhat unusual when compared to the standard (i.e. non-iterated) experiments that we often see in psychology. Just to give one example, whereas in most experiments the performance of each individual participant on a given task is measured as a data point, here a data point corresponds to an individual language that has been passed between many participants. This makes these kinds of experiments relatively expensive to conduct: in order to demonstrate significance we must recruit many more participants per condition. Although I hope to show that this endeavour is worthwhile, and that human iterated language learning experiments are an invaluable method for gaining insight into how the very act of learning affects the structure of systems for future learners, I also hope that this work goes on to inspire more realistic computational and mathematical models of the process.

In short, the theory that I will be testing is that language adapts. More specifically, language adapts to suit the conditions under which it is transmitted. As with any kind of problem to be solved, there are often multiple solutions. Variation in language arises because there are many ways in which a language can be structured, all of which are equally well adapted to the task of being transmitted. The fact that languages are culturally evolving systems can thus explain why they are open-ended *and* variable. Notice that so far we have not made any mention of communication. This is deliberate. Although language gives all the appearance of having been designed for communication (e.g. Pinker, 2003), I will take the somewhat unusual approach of investigating the extent to which linguistic structures that are communicatively useful could have arisen without it. If it turns out that communication is not required in my experiments to get these structures to emerge, this does not prove that language evolution necessarily happened in that manner. However, it would require us to think more deeply about the possibility.

Another issue I will be exploring which tangentially relates to this concerns the nature of the mechanisms responsible for evolutionary change. Language, unlike the Scott Monument, was designed without a designer. Although in a sense it was created by intentional beings, it was not the intention of those beings to create it (Keller, 1994). Croft (2000) agrees, and makes a useful distinction between what he calls intentional changes (where a speaker has some other goal in language use in mind, and produces some unforeseen innovation along the way) and nonintentional changes (where a speaker has no goal in mind at all but introduces a change as a consequence of the act of production or comprehension itself). I will show some examples of empirical studies which, by this definition, explore the intentional design of communication systems in the laboratory (§3.2.4), but argue that we also need to investigate the possibility that structural features of language could also have emerged through more nonintentional processes.

To summarise, this work addresses the following questions:

1. Why is language structured the way it is and not some other way?
2. How does the process of cultural transmission give rise to language structure?
3. Can features of language structure which appear to be designed for communication evolve in the absence of a) actual communication, and b) intentional design?

1.3 Experiments

The basic methodology of the experiments is based on an agent-based computational simulation of cultural transmission, known as the iterated learning model (Kirby & Hurford, 2002; Brighton *et al.*, 2005), and involves the transmission of small artificial languages between human learners. Participants are recruited and told they must learn how to speak a newly discovered alien language. During training, they are shown images (meanings) of different coloured shapes engaged in some kind of motion, along with a written description (signals) showing how the alien would refer to that particular image. After training, their knowledge of the alien language is tested by showing them each meaning in turn and asking for the correct signal. Whatever is produced as output in this final test then becomes the new training data for the next participant. This process iterates to form a linear diffusion chain of learners, each of whom have unknowingly acquired their language from the previous participant.

The first experiment looks at what happens when learners are only given partial access to signals and meanings during learning, whereas the second experiment looks at what happens when this restriction is lifted. The third experiment explores what happens to the languages when we make an invisible modification to the process of transmission, such that only unambiguous signals get transmitted to future participants. The fourth study builds on this, and looks to see what effect increasing the amount of training has. The fifth study is somewhat different to the previous four in that it does not involve the transmission of meaning-signal pairs at all. In this final study, we try to get a better look at how sequence learning constraints may influence things, by focusing on how signals evolve in the absence of any meanings.

Each experiment can be seen to stand alone, operating to investigate its own particular hypotheses. However, they have also been designed with specific contrasts in mind. Within the four main experiments, three conditions are examined. Experiments I and II differ only in terms of whether or not participants have access to the full language during learning; Experiments I and III differ only in terms of whether or not unambiguous signals get passed on to learners; and finally, Experiments III and IV differ only in terms of the amount of exposure to training items that learners receive.

1.4 Thesis Road-Map

In ***Chapter Two*** we take a closer look at language and the key approaches that have been taken to explain its emergence. In particular, it presents an account of iterated learning -- the process of cultural transmission at work in language -- and summarises the key findings to have emerged from computational and mathematical models of the process in language, and in different domains. It will then move on to explore some of the literature on cultural evolution more generally. In particular, it looks at some of the more influential theoretical accounts, the effect that the direction of cultural transmission has, and finally, reviews some of the experiments that have been done to explore the mechanisms and dynamics of cultural evolution in both humans and non-humans.

Chapter Three also reviews literature, but this time focuses specifically on attempts to empirically investigate language origins in humans. Its main purpose is to motivate the design of the current methodology. It reviews the current approaches to explaining language emergence, both inside and outside of the laboratory, and argues that although language arises through the actions of intentional beings, it has not been intentionally designed or created in any way. In order to isolate and better understand this unintentional aspect of language emergence, we need an experiment design that does not involve intentional communication between participants. The details of this design are then laid out ahead of the actual experiments themselves.

The first set of results are reported in ***Chapter Four***. In this chapter we look in more detail at one of the key parameters from the iterated learning models discussed in Chapter 2 -- the transmission bottleneck. The first and second experiments test out predictions made by the computer simulations, and find that although the main findings associated with iterated learning studies -- that languages evolve to become easier to learn and more highly structured -- are replicated in human learners, there are some interesting differences.

These differences are further explored in ***Chapter Five***, where we focus in on the natural tension that exists between learnability and expressivity. The third and fourth experiments are outlined here, showing that when we add in a pressure for greater expression of the meaning-space, we start to see signs of compositionality emerging in the languages. Techniques are introduced which allow us to precisely quantify the emergence of this compositional structure, and which enable us to see how cultural transmission amplifies local structural regularities in the input and allows them to accumulate over time.

Chapter Six takes a very different approach, and asks the question of whether we can try to isolate the effects of some of the learning biases that are at work in the minds of participants. It introduces several modifications to the experimental methodology designed to eliminate other biases, one of which entails the complete removal of meanings. The results of this study show that even when there are no pressures upon signals to adapt to express structured meanings, signals nevertheless begin to show signs of structure as a result of the sequence memory constraints of the learners.

Finally, ***Chapter Seven*** returns to some of the key themes expressed throughout the thesis and attempts to link them to some of the wider issues within the field of evolutionary linguistics. It presents a brief summary of the major points emerging from the five studies, and contends that the key contribution of all this work lies not just in the lessons we have learned from the various experimental manipulations that have been explored, but in the development of the experimental methodology itself. It stresses the significance of cultural transmission in the process of language evolution, and suggests that the next challenge facing the field lies in explaining where the mechanisms underlying iterated learning come from.

Chapter Two

Language and Cultural Evolution

The first chapter has set the scene for thinking about language as an evolving system in its own right, and given an overview of the general direction of the rest of this thesis. The rest of this chapter outlines in more detail some of the reasons we might be interested in studying language, and in particular, the origins of language, before moving on to explore some of the ways in which the topic of language evolution has been approached recently. It briefly introduces the reader to the iterated learning framework, which forms the theoretical backbone of the thesis, and then moves on to explore work undertaken in the field of cultural evolution more generally. It presents a very brief overview of the main theoretical approaches, before finally exploring some of the experimental work undertaken using both human and non-human participants.

2.1 Some Facts about Language

Language defines us as a species

Language is often credited with being the behavioural trait that defines us as a species. There are perhaps two main reasons for this assertion. The first is to do with the special role that it seems to play in our lives. We use language. A lot. And not just for simple communication. We use it during cognition (Dennett, 1995; Jackendoff, 1996; Clark, 1998), for co-ordinating joint actions (Clark, 1996), when constructing a theory of mind about others (Tomasello, 1999), for maintaining social bonds (Dunbar, 1996), and for categorising objects in our world (Vygotsky, 1962; Lupyan *et al.*, 2007), to name just a few. If other species are using their communication systems for all of these extra purposes, there is surprisingly little evidence for it.

The second reason for claiming our linguistic abilities separate us from other creatures relates to the properties of language itself. We are not just different from other animals in how we use our communication system, we differ in how that system itself works. If we focus on the features that all human languages share with one another, and then look to find correlates to those features in other communication systems in nature, we can identify the similarities and differences. Although many of these proposed 'design features' are shared with other species, some appear genuinely unique to humans (Hockett, 1958; Hockett & Altmann, 1968).

Language is open-ended and variable

Of all the features claimed to be universal and unique to language, two seem particularly striking: unlike other natural communication systems, human languages are, (a) open-ended, and (b) highly variable. It turns out that even explaining these two traits presents us with some interesting evolutionary problems. If we begin with the open-endedness of language, we can easily understand how a system which is capable of expressing an unlimited set of concepts might be useful. The ability to communicate a novel thought, in a novel context, perhaps to a novel interlocutor, using a novel packaging of signals, is not to be sniffed at. If I were to wish you 'sweet elbowy lamb dreams', you might think I was behaving strangely but you would nevertheless understand the basic message. This is in spite of the fact that you could not possibly have heard that particular sentence before, may never have met me, and are entirely removed from the context of the utterance[1].

The open-endedness of human language is a result of the way that it is structured. In particular, all languages exhibit **duality of patterning** (meaningful units are created by the reuse and recombination of smaller meaningless units), **compositionality** (more complex meanings are created by the structured ordering of meaningful units), **systematicity** (there is a structure-preserving relationship between signals and meanings), **regularity** (relationships between signals and meanings, and other structures at higher levels, are expressed reliably and unambiguously) and possibly **recursion** (rules of language can be self-referencing, allowing for complex embedding and hierarchical ordering of clauses)[2]. These structural properties are universal, perhaps even definitional, of language. Without the property of compositionality for instance, we could not interpret the meaning of novel sequential arrangements of words even if those words were familiar, and if the relationship between signals and meanings were unsystematic and irregular, we could not make generalisations over utterances and apply them to new situations.

These structural properties are more than just a bag of neat linguistic tricks - they are integral to explaining how humans have managed to survive in almost every habitat on earth (Hurford *et al.,* 1998; Fitch, 2010), and build the technology required to escape the confines of our planet (Kirby & Christiansen, 2003). In contrast, even though many animals are capable of complex thought and reasoning, they are still restricted to more limited domains of expression (Hurford, 2007; Fitch, 2010). So whilst Vervet monkeys can famously differentiate between different types of predators and make alarm calls accordingly (e.g. Cheney & Seyfarth, 1990), they

[1] For the curious, the context here is craving lamb *codillo* before bed time, and sadly nothing to do with falling asleep trying to count sheep who are jostling each other.

[2] The issue of whether recursion is present in all languages is a contentious one. See Everett (2005), Parker (2006), and Luuk & Luuk (2010) for the argument against, and Fitch *et al.* (2005), and Hauser *et al.* (2002) for the argument for.

cannot create novel alarm calls for new predators (even though they can perceive them), or use their existing calls for novel purposes (other than triggering a flight response). Given the obvious utility and adaptive value of a system capable of unlimited expression, why has this trait not evolved in species other than our own?

Moving onto the inherent variation in language, we find that it goes beyond the fact that there are some 6,000 or so different languages existing in the world today[3]. Variation also exists within the same language community, both synchronically in the form of different dialects, and diachronically in the form of different historical variants. Even if we focus down to the level of an individual speaker, we find immense variation in the choice of particular words, phrases, intonation patterns and pronunciation of phonemes, based on any combination of social, contextual, emotional and articulatory factors operating at any given moment. In short, variation exists at all levels of organisation within language, across languages, and at both the population and individual level (Evans & Levinson, 2009).

Again this presents us with a problem: having this much variation in language entails that language must be learnt, and biologically speaking, learning is a costly process. Indeed, we see that while many species have offspring who are capable of walking and catching themselves a good meal within minutes of being born, human infants are entirely dependent on their parents for survival, and do the majority of their development outside of the womb. Before they start learning how to walk, babies are learning how to talk, devoting the majority of their cognitive resources to this one task. Obviously, we know that learning, and social learning in particular, brings other benefits that must outweigh these costs (Barrett *et al.*, 2002). However, it

---

[3] Although this in itself is highly unusual if we are drawing comparisons with animal communication systems, which tend to be innate and therefore uniform across all members of the same species (Evans & Levinson, 2009). A rare exception to this pattern are the systems of some species of song-birds, seals and cetaceans who learn their songs culturally from conspecifics, which often results in geographical variation in the structures of songs sung by members of the same species (Marler & Tamura, 1962; Doupe & Kuhl, 1999; Rendell & Whitehead, 2005).

has been suggested that humans are doing something special when it comes to learning language.

Language acquisition is automatic

Although language is an incredibly complex system, with many intricate context-dependent rules and exceptions, infants seem to acquire it effortlessly. In fact, by the age of four, all healthy children will have mastered the basic structures of syntax (Bates *et al.*, 2003), and all without taking advantage of any direct instruction or correction by their caregivers (Hirsh-Pasek *et al.*, 1984). There are many well-attested developmental patterns within language (Pinker, 1994). For instance, it has been shown that the order that children acquire certain bound morphemes in English is the same across learners, and that this is unrelated to the frequency with which those morphemes appear in the speech of caregivers (Brown, 1973; Slobin, 1982). Similar findings are found in the development of phonological (Locke, 1983), syntactic (Ingram, 1989) and semantic (Johnston, 1985) aspects of language as well.

This is interesting, and when combined with data from cross-linguistic studies, which indicate broadly uniform developmental sequences across different languages and cultures (Brown & Hanlon, 1970; Slobin, 1982), suggests that these patterns cannot be explained by the linguistic environment alone. However that is not to say that the linguistic environment is not important. There is strong evidence for a critical period in language acquisition -- a certain 'window of opportunity' where learning language is possible, thought to last between infancy and puberty (Lenneberg, 1967). If learners are deprived of input during this time period, they will not go on to develop full linguistic competence (Curtiss, 1977; Skuse, 1984).

These three facts combined -- that language acquisition proceeds reliably, exhibits universal developmental patterns, and that there is a critical period for it -- has led most, if not all, researchers to the conclusion that there is some innate component constraining the acquisition process, even if it is not specific to language (e.g. Elman *et al.,* 1996). Given the fact that attempts to teach human language to non-humans have all failed (Fitch, 2010), it also seems reasonable to conclude that this innate component must also be somehow specific to humans, even if it is more the result of quantitative rather than qualitative differences in cognition (Hauser & Fitch, 2003; Hurford, 2004). However, as the qualifications in the previous two sentences suggest, there is still scope for much disagreement as to what this innate contribution might be, what the role of learning is, and what kind of evolutionary mechanism(s) are responsible for it.

2.2 Key Approaches to Language Evolution

Before we can explain how language evolves, we need to be able to explain what it is and how it works. Broadly speaking, two different approaches have been taken to explain language, and as a result, its evolution. Each differs to the extent that it sees language as being the end-product of specialised cognitive machinery, and also to the importance to which it ascribes processes of cultural transmission. This section explore these two stances in more detail.

2.2.1 The Direct Appeal to Biology

The first approach, taken by some to be the 'standard' or orthodox evolutionary view (e.g. Kirby *et al.,* 2008b), is to suggest that the structure of human language can be explained by a direct appeal to biology. At its heart, this approach rests on the claims originating with the linguist Noam Chomsky concerning how children acquire language. According to Chomsky (1959; 1965; 1980), language learning is constrained by an innately-specified language acquisition device (LAD) which shapes the kinds of hypotheses the child is willing to entertain about language, and ultimately guides them to the correct grammar. As the LAD is as much a part of our biology as, for instance, the human eye, we can account for its evolution in the same way: by treating it as an adaptation (in the case of language, for communicating propositional meanings), brought about via processes of natural selection (Pinker & Bloom, 1990; Pinker, 2003).

The Nativist Position

This idea encapsulates a very specific notion of innateness. This is reflected in the terms used by theorists to describe what is going on. Humans, it is said, are in possession of an 'innate module' (Fodor, 1983), a 'language instinct' (Pinker, 1994) or a 'faculty of language' (Chomsky, 2002). What unites all subscribers to this kind of nativist view is a single shared tenet: that the primary determinant of language acquisition is a body of innate knowledge specifically pertaining to language[4]. Under this view, universal structural properties of language are seen as the direct expression of the genes - they appear in language because of constraints from our innate learning mechanisms (Fig 2.1). Whilst languages themselves may vary, the fundamental ways in which languages are organised do not, because they are genetically determined. Although nativists recognise the crucial role that linguistic input plays in triggering this process and interacting with the information held within the LAD, what ultimately causes language to exhibit the hallmark structural features that it does can only be explained by understanding what children are born with and bring to the task of learning.

Why might we want to make such a strong (and as we shall see later, controversial) claim? The first reason is that an innate LAD can help us explain the reliability, uniformity and universality of acquisition discussed earlier. Language seems to

[4] Chomsky has recently clarified his position with regards to the faculty of language, distinguishing between the faculty of language in the broad sense (FLB), which contains cognitive mechanisms that are either not specific to humans or not specific to language, and the faculty of language in the narrow sense (FLN), which contains only cognitive mechanisms specific to both humans and language (Hauser, Chomsky & Fitch, 2002). I am only referring to FLN here, although it is important to remember that the contents of FLB can also be explained by a direct appeal to biology.

unfold in predictable stages, much as other instinctual behaviours in the animal kingdom do. The second is more technical, and relates to *the argument from the poverty of stimulus* (Chomsky, 1965; Wexler, 1991). Nativists have traditionally approached the issue of language learning as a problem of grammar induction: given a set of data, the child's learning task amounts to reconstructing the grammar responsible for generating that data. The problem with this however, is that the stimulus data a child observes will underdetermine this grammar every time. In other words, there is not enough evidence in the primary linguistic data available to children to allow them to induce the correct grammar with any degree of certainty.

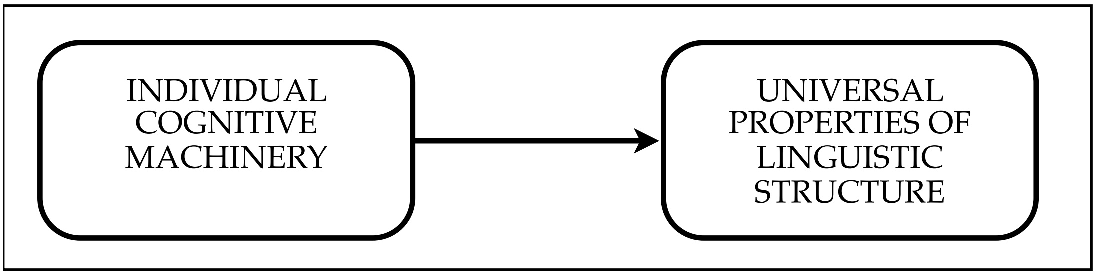


Fig 2.1: The nativist position with regards to explaining the appearance of universal properties of linguistic structure. It is claimed that there is a direct link between an individual's cognitive machinery and structural patterns seen in the world's languages. Based on Kirby, Smith & Cornish, (2008).

Pullum & Scholz (2002) surveyed the language acquisition literature and compiled a list of six frequent claims made by researchers concerning the properties of the child's learning environment (see Fig 2.2 below). The basic facts here are not in dispute, however there are still significant disagreements as to how we should interpret them. The property of POSITIVITY has perhaps created the most debate. The data the child is exposed to is not only finite, idiosyncratic and incomplete, but it also consists only of positive examples of legitimate sentences. This makes it compatible with an infinite number of different hypothetical grammars, which in turn makes the task of converging upon the single correct grammar that produces that data, and only that data, akin to finding the needle in the proverbial haystack

(Gold, 1967; Hendriks, 2000). Without evidence of what is *ungrammatical* in the language, how is the child supposed to discard incorrect hypotheses?

Obviously, if children can only entertain hypotheses about grammar that are licensed a priori by some innate and specialised language acquisition mechanism, then this problem is solved. In this way researchers can also address the continuity problem: not only can they account for how language is acquired (children are biologically constrained to only look for certain types of grammar), but also why it is that human languages occupy just a small subset of those that are logically possible (languages all have similar underlying structural properties because they are created by humans who all have the same set of biological constraints) (Crain & Pietrosky, 2001).

Properties of the child's environment

a. INGRATITUDE: Children are not specifically or directly rewarded for their advances in language learning.
b. FINITENESS: Children's data-exposure histories are purely finite.
c. IDIOSYNCRASY: Children's data-exposure histories are highly diverse.
d. INCOMPLETENESS: Children's data-exposure histories are incomplete (there are many sentences they never hear).
e. POSITIVITY: Children's data-exposure histories are solely positive (they are not given negative data, i.e. details of what is *un*grammatical).
f. DEGENERACY: Children's data-exposure histories include numerous errors (slips of the tongue, false starts, etc.).

Figure 2.2: A list of claims frequently made by language acquisition researchers concerning the properties of the child's learning environment. These claims are not disputed in themselves, but their interpretations are still the subject of much discussion. Taken from Pullum & Scholz (2002:13).

Language: a Naturally Selected Biological Adaptation

Pinker & Bloom (1990) have argued that we can explain the evolution of language in the same way we would explain the evolution of any organ in the body: as an adaptation. They go on to argue that the evolutionary process responsible for this adaptation must be natural selection.

> "Evolutionary theory offers clear criteria for when a trait should be attributed to natural selection: complex design for some function, and the absence of alternative processes capable of explaining such complexity. Human language meets these criteria." (Pinker & Bloom, 1990:707)

Language undoubtedly holds the appearance of design. For Pinker & Bloom, this design clearly relates to the function of communicating propositions through a serial transmission channel. Obviously this process did not happen overnight, and neither did language as we know it spring out fully formed in one go. At some point we must explain how language arose out of non-language. According to Pinker & Bloom (1990), and later Pinker (2003), natural selection is a viable solution to this problem of emergence as long as any small ability to communicate was slightly advantageous. In the same way that the eye developed gradually -- at first just as a few cells capable of perceiving light and dark that might have allowed an organism to perceive when a predator was close-by, before later being able to differentiate separate frequencies of light, which may have helped an organism to avoid poisonous foods -- so too did language evolve. In increments.

This theory still contends that the proximate cause of the structural properties we see in languages in the world lies in specialised cognitive machinery, but adds to this the claim that what ultimately causes it is biological evolution under natural selection for communication. This yields the following set of relationships between language, learning, and evolution (Fig 2.3). As we can see, this modifies the picture in Fig 2.1 only slightly.

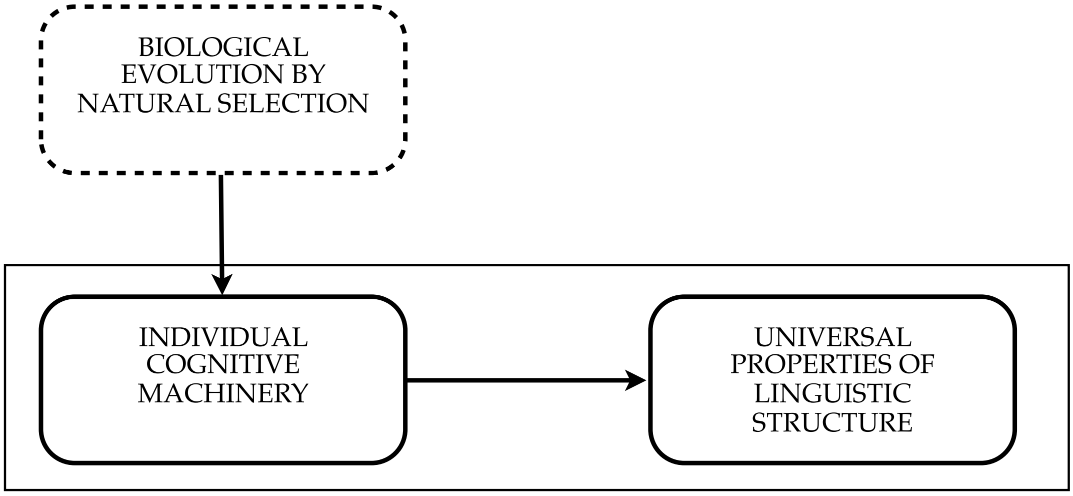


Fig. 2.3: The orthodox evolutionary view in full. Universal properties of linguistic structure are directly caused by the nature of our individual cognitive machinery. This machinery in turn, is the result of biological evolution, which is under natural selection for enhanced communication. Re-drawn from Kirby, Smith & Cornish (2008).

There are basically three ways in which this idea has been challenged. The first is in terms of whether language really is an adaptation *for communication* or not. Dunbar (1996) and Miller (2000) both support the adaptationist stance, but disagree as to what the primary function of language was when it evolved. There are two options here. Either language evolved for something other than communication, and is still used for that other function, or it could in fact be an exaptation: an adaptation for something else that has since been 'borrowed' and further tweaked to suit a new purpose (Gould & Vrba, 1982).

What other purpose could language (or its precursor) have served? Dunbar (1996) sees language primarily as a method for instilling social bonds, what he refers to as 'social grooming'. The argument here is that as group sizes increased in our hominid ancestors, one-on-one manual grooming, a main-stay in primate social interactions, became impractical. Vocal gestures, unlike physical gestures, can proceed in a one-to-many fashion. As such, language might have evolved to replace manual grooming and maintain social contracts between individuals. Miller (2000) on the other hand, agrees that language is a biological adaptation, but disagrees both with the function and the evolutionary mechanism responsible. For him, the pre-cursor to language was an integral part of the courtship process between early humans, and therefore language is at least partially the result of a process of sexual selection. Both of these accounts have been criticised individually[5], but a common complaint with them both is that neither explains exactly why features of language seem so well designed for communication and not any of the other alternative functions proposed (e.g. Pinker, 2003). They do, however, serve to highlight the range of alternatives that could be considered even when we adhere to the simple idea that language need only be understood in biological and adaptationist terms[6].

Another way that Pinker & Bloom's idea has been challenged relates to the relationship between the innate learning mechanisms and the properties of language shown in Figs 2.1 and 2.3. The orthodox account assumes that the link between, on the one hand, the cognitive machinery in an individual learner's brains, and on the other, the behaviour that that machinery manifests at the population level, is a direct and transparent one. But what if it is not? Kirby (1999:19-20) refers to this issue as the *problem of linkage:*

> "The innatist approach links universals to acquisition, so that constraints on cross-linguistic variation are the *direct consequence* of constraints on the acquisition (and mental representation) of language.[...]To be completely explicit, we can formulate the following problem:

---

[5] Nakamura (2000) has called into question Dunbar's (1996) assertions that verbal grooming is inherently more efficient than manual alternatives, whereas Miller's (2000) claims have been challenged on the grounds that it predicts elaborate but ultimately meaningless signalling displays - not compositional syntax (Pinker, 2003).

[6] There have been non-adaptationist theories put forward to explain language evolution as well -- most famously Chomsky (1988), Piatelli-Palmarini (1989) and Piattelli-Palmarini & Uriagereka (2004). All of these theories adhere to the nativist position, but caution against assuming language was naturally selected 'for' anything. A full discussion of these other theories is outwith the scope of this review, although see Gould (1997) for a general discussion on the merits of non-adaptationist explanations for human evolution.

> **The problem of linkage.** Given a set of observed constraints on cross-linguistic variation, and a corresponding pattern of functional preference, an explanation of this fit will solve the problem: how does the latter give rise to the former?" (emphasis original)

In other words, we need to be able to account for exactly *how* patterns of neural activity actually wind up as patterns of linguistic behaviour (Kirby, 1999; Kirby *et al.,* 2004)[7].

The final criticism is also related to this. As Pinker & Bloom state themselves, the compulsion to accept an explanation involving natural selection holds only as long as there are not, in fact, 'alternative processes' that could explain the appearance of design. As Kirby (2000) claims, and indeed, the next section will discuss, there is an alternative process capable of explaining the appearance of design in language -- and it also has the added advantage of solving the problem of linkage for us.

2.2.2 Language as a Complex Adaptive System

The nativist explanation of language -- and more recently language origins -- has been the dominant approach in both linguistics and cognitive science for many years. However, it is not the only approach. The poverty of stimulus argument, which forms the cornerstone for acceptance or rejection of the proposal, has been increasingly under attack, with neither side managing to produce conclusive evidence for or against (Pullum & Scholz, 2002). Some claim that the poverty of stimulus argument is tautologous, and question whether we can view language learning as a strictly-rational process of grammar induction at all (e.g. Tomasello, 1995; Hendriks, 2000; Tomasello, 2004), whilst others argue that nativists are over-stating the paucity somewhat, or question the claims that general-purpose learning

---

[7] Strictly speaking this should be considered a critique of the nativist position in general rather than one levied specifically at Pinker & Bloom (1990).

mechanisms really do all that badly with sparse input data (Marcus, 1993; Elman *et al.*, 1996; Cowie, 1999; Gomez & Gerken, 2000).

Given that we might also have doubts concerning the problem of linkage between the structure of a language learner's cognitive machinery and linguistic behaviour at the population level (§2.2.1), what are we left with? We must still account for the facts we have learnt about language acquisition, comparative studies of animal communication systems, and the underlying similarities between different languages. The alternative suggestion is to rethink what we mean by the term 'innate' (c.f. Elman *et al.*, 1996).

Clearly there is something special about human biology. There is a good deal of evidence to suggest that we have undergone many physiological changes or pre-adaptations for language, most notably our transition to bipedalism (which allowed for greater breath control), and alterations to our vocal tract and perceptual systems (Hurford, 2003). However, increasingly, and for the reasons specified above, researchers have been reconsidering whether the cognitive mechanisms that underlie language learning, processing, and use, really have to have been specially developed for language.

Many of those who subscribe to this belief take a complex adaptive systems (CAS) view of language. That is, rather than seeing language as the sole result of a psychological process ongoing within the individual, they see language as an emergent phenomenon, arising as the result of a series of many local interactions between speakers that give rise to more complex behaviours at higher levels (Gell-Mann, 1992; Holland, 1995; Hashimoto, 2002; Brighton *et al.*, 2005; Christiansen & Chater, 2008; Beckner *et al.*, 2009). This approach tends to rely less on the notion of cognitive mechanisms specific to language, and recognises that languages themselves are adaptive systems capable of undergoing their own form of (cultural) evolution. This next section explains this position in more detail.

The field of linguistics has long recognised that languages are historical entities that change over time. However, attempts to integrate diachronic linguistics with more synchronic approaches have not always been successful. One of the advantages of taking a CAS view of language, is that it can lead to a natural coming together of these two sides of the same coin. Essentially, we can see language as the result of the interactions between three different adaptive systems, each of which operates over a very different time-scale (Kirby & Hurford, 2002). Figure 2.4 shows some of the possible interactions between these different systems.

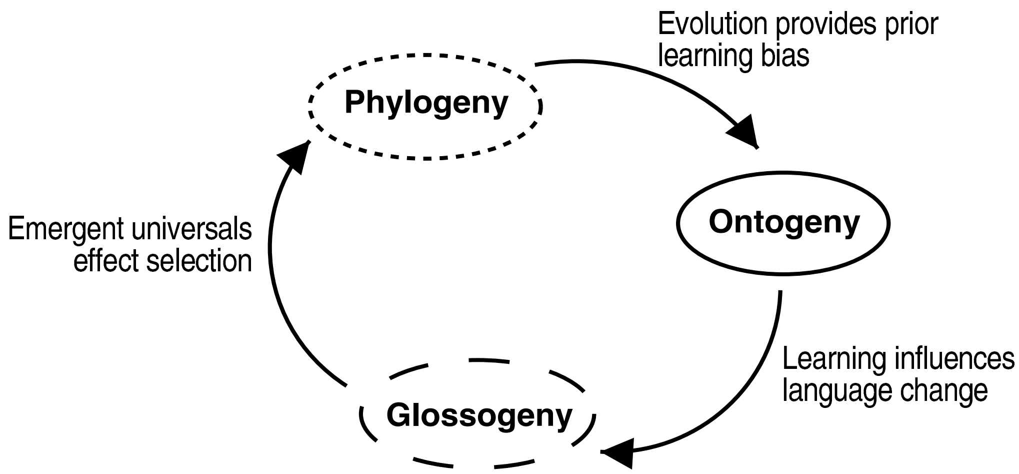


Figure 2.4: Language is the result of three complex adaptive systems. According to this perspective, interactions between the different systems are important. A few of these possible interactions are shown here. Biological evolution gives rise to phylogenetic changes which provide a platform for learning; this creates a set of learning biases which in turn largely influence what can be acquired ontogenetically; this in turn affects which features of languages persist culturally, and what kinds of glossogenetic changes occur; these emergent structural features finally feed back into biology, by influencing the selection pressures on the evolving speakers of that language. Taken from Kirby & Hurford (2002).

At one level, we have **phylogeny**, which relates to the biological evolution of the learning and processing mechanisms (general, or otherwise) used for language. This system operates over the time-scale of the evolution of the species, and provides learning biases which go on to interact with the next system, **ontogeny.** Ontogeny relates to the development of the capacity for language within an individual -- in other words, language acquisition. Learning is itself an adaptive process, with operates over the life-time of the individual learner. It is influenced by biological learning biases, but also goes on to influence our third system, dubbed **glossogeny** (Hurford, 1990). Glossogeny is a process relating to the way that languages themselves adapt and change over a historical time-scale, which we can think of as equivalent to the 'lifetime' of a specific language. Adaptive changes at this level are influenced as a result of learning undertaken not just by one individual, but by many. The resulting structures that emerge go on to further influence the evolution, by providing selection pressures for learning biases that better accommodate these emergent features of language.

Cultural Transmission: The Missing Link?

We learn language by observing the linguistic data produced by others. This alone is enough to make language a cultural system. The real question is not about whether this cultural system exists, but about whether it contributes anything to the process of linguistic emergence. In other words, does cultural transmission actually change the story presented in Fig. 2.3 in any significant way? As Kirby et al. (2008b) acknowledge, it could well be the case that all cultural transmission does is act as a passive conduit, linking the cognitive learning biases in our heads to the linguistic structures in the world as a mere intermediary step. However, it could also be the case that processes of social interaction and cultural evolution actively generate structure, and provide us with a way to bridge the gap between individual-level cognition and population-level behaviour and solve the problem of linkage once and for all (Kirby, 1999).

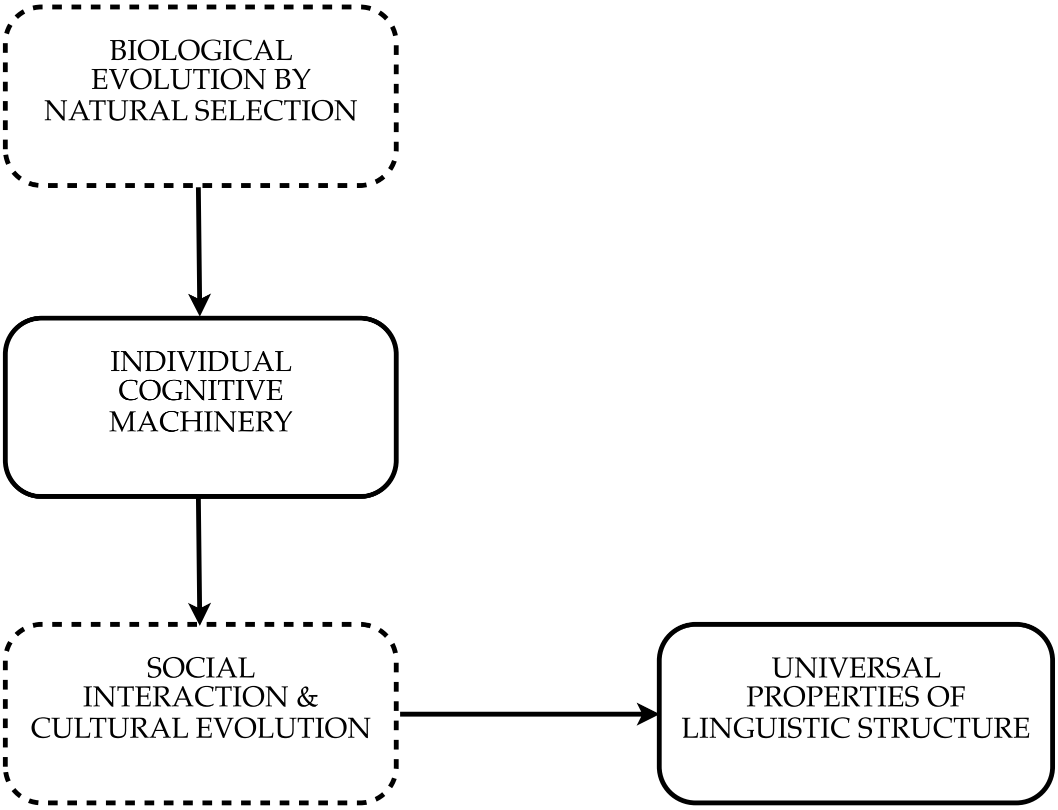


Figure 2.5: Solving the problem of linkage. Processes of social interaction and cultural evolution are thought to have a constructive role to play in explaining how we get from learning biases in individual brains, to linguistic behaviour in populations. Redrawn from Kirby, Smith & Cornish (2008) with permission.

Fig. 2.5 shows how this fits in with our previous diagrams. Here we see that the link between our cognitive learning biases and universal properties of linguistic structure are mediated by processes of social interaction and cultural evolution. By studying these processes in more detail we will gain insight as to what, if anything, they can contribute to our understanding of where structure in language comes from. One type of cultural transmission mechanism in particular is thought to be capable of bridging this gap between individual minds and the behaviour of populations. The next section examines this mechanism in more detail.

2.3 Iterated Learning: A Mechanism of Cultural Transmission

There is something special about the way in which language is acquired. Language learning involves learners learning from other learners. More formally, this process has been referred to as **iterated learning**.

> "Iterated learning is a process in which an individual acquires a behavior by observing a similar behavior in another individual who acquired it in the same way" (Kirby, Cornish, & Smith, 2008: 10681)

It is important to note at this point that iterated learning is a domain-general process and not unique to language -- it can apply to other domains of learned behaviour (Brighton, 2003), and is not a process which operates exclusively in humans (Feher *et al.*, 2009). It also makes no specific claims about the particular population structure learners are configured in -- it applies equally to inter-generational and intra-generational interactions between learners (see §2.4.3 and §3.4.1 for more details). Finally, it should be remembered that it is just one of just a number of mechanisms of cultural transmission, such as imitation or teaching[8], albeit the one most relevant to language.

The fact that there is feedback or interaction between the learner and what is being learned does make it different from many other types of observational learning that we engage in however. We can think about this in the following way. Learning a language is not like learning how physical objects move in the world. This is because the properties of the aspect of the physical world that we learn about when we learn how objects move, have been entirely constrained by processes external to

8 Like both of these mechanisms of cultural transmission, and unlike for example, emulation or stimulus enhancement, we will see over the course of this thesis that iterated learning is also capable of giving rise to cumulative cultural evolution. For more information on the differences between imitation, teaching, stimulus enhancement and emulation in more detail, see Tomasello (1999). See also §2.4.4 for more discussion on the cumulative nature of human culture.

our cognitive system. The properties of language on the other hand, are actually determined by the learning efforts of previous learners -- which, to the extent that learners have similar learning biases, means that an initial intuition that a learner might have about how a particular linguistic structure works will most likely be correct (Christiansen & Chater, 2008). Interestingly, arguments of this type turn poverty of stimulus claims on their head. As Zuidema (2003:58) puts it:

> "[L]earners are only presented with targets that other learners have been able to learn. [...] The poverty of the stimulus is now no longer a problem; instead, the ancestors' poverty is the solution to the child's."

This is just one of the interesting implications that studies of iterated learning as a cultural transmission mechanism reveal. The rest of this section focuses on the conceptual framework for understanding iterated language learning more specifically, and then explores some of the main findings to have emerged from research into iterated learning using computational and mathematical models.

Iterated Language Learning

How does iterated learning apply to language? It was in fact Chomsky (1986) who argued that language exists in roughly two forms – **E-Language** ('external' language, represented in the world by actual utterances, and a property of populations of speakers/hearers) and **I-Language** ('internal' language, represented in the minds of speakers as a pattern of neural connections, and a property of an individual speaker/hearer). Language induction involves the transformation of E-Language into I-Language, as each learner induces their own mental representations of language on the basis of exposure to the ambient language surrounding them. On the other hand, language production involves the reverse mapping, as agents use their internal representations to create new utterances, which creates the external language for the next generation to learn from.

When this process of induction and production iterates across several learners, each learning from the output of the previous generation, it becomes an (iterated learning) model of language evolution (Hurford, 2000). This process is schematised in Fig. 2.6 below. Because of the way that this framework attempts to explicitly understand the link between individual learners and properties of language, it directly speaks to the issue of the problem of linkage (Kirby *et al.*, 2004).

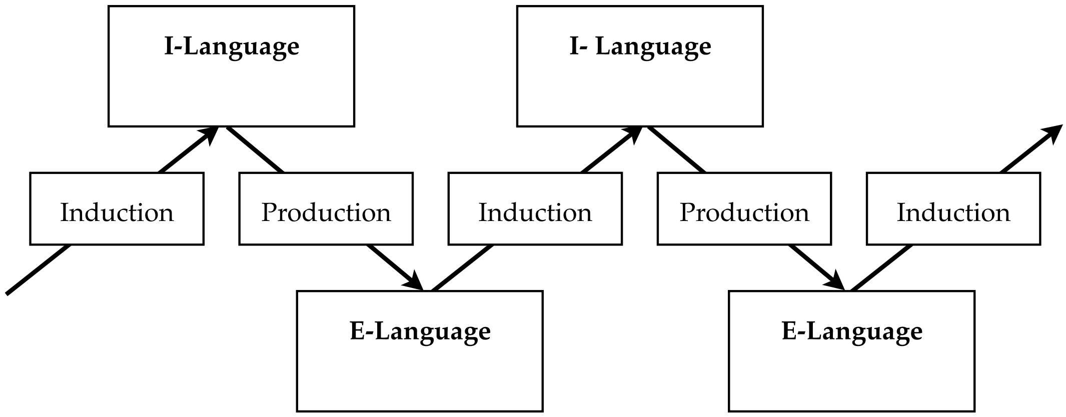


Figure 2.6: The transformation of I-Language into E-Language over successive generations or interactions. Each learner induces an internal mental representation of language (I-Language) by observing utterances that are publicly represented in the external world (E-Language). Learners then become speakers, and produce new utterances, possibly changing the content of E-Language in some way. When this process iterates it becomes an evolutionary system. Redrawn from Kirby (2001) with permission.

In this account, the role of previous language users is crucial. The process of iterated learning is imperfect: as information is transformed between the different domains during induction and production, there is a chance that small linguistic changes will be introduced[9]. These changes are not simply errors that the next learner will ignore or correct. In most cases these errors will be indistinguishable from non-errors, and will go on to influence the linguistic system of future learners accordingly. As

[9] See Hoefler (2009) for an interesting discussion about loci for the introduction of changes in the transmission cycle in more detail.

Brighton (2003:35) puts it: “language reflects the accumulated residue of the effects of learning and production of preceding agents.”

The Iterated Learning Model: Some Examples

As stated in the previous chapter, the majority of work investigating the process of iterated learning, particularly in relation to language, has focused on building computational and mathematical models of it. Many of these computational simulations are agent-based models, which explicitly attempt to simulate both the cognitive processes of individual agents, as well as learning interactions between different agents. What differentiates these models from others investigating language evolution is the fact that there is no genetic evolution involved, and agents are not rewarded in any way for successful communication.

A typical simulation consists of one or more **learning agents**, one or more **teaching agents**, a **meaning space** consisting of a shared set of concepts an agent can talk about (usually represented by a vector, real number, or a logical proposition), and a **signal space** which is initially empty. An agent is selected to be a teacher and randomly chooses a sub-set of meanings to express from the meaning space. If they do not already have a signal for a given meaning, the agent – who is equipped with the ability to produce a string at random – will create one. These signal-meaning pairs produced by the teacher are then given as training data to the next learner agent, which uses this to develop its own representation of the data using some kind of induction mechanism, and the cycle repeats.

Over the years, various parameters have been explored: different types of production and induction mechanisms (Batali, 1998; Kirby, 2000; Brighton, 2002; Tonkes & Wiles, 2002), different structures and sizes of meaning-space (Batali, 1998; Kirby, 2002b; Teal & Taylor, 2000; Zuidema, 2003; Kirby, 2007), and different population structures (Batali, 1998; Livingstone & Fyfe, 1999; Kirby, 2000; Batali, 2002; Smith & Hurford, 2003; Vogt, 2007) to name just a few. One of the most robust findings however, seen in every condition tested so far, is that the resulting languages created by the agents become easier to learn over time. The key parameter responsible for this result is known as the **transmission bottleneck**. Deacon (1997:110) was one of the first to put words to this phenomenon in recent literature:

> "Languages are social and cultural entities that have evolved with respect to the forces of selection imposed by human users. The structure of a language is under intense selection because in its reproduction from generation to generation, it must pass through a narrow bottleneck: children's minds."

In order to survive to be present in the external pool of language – or in other words – in order to be transmitted and stand a chance of becoming part of I-Language in the future, it must be learnable (by humans or simulated agents). There are several ways in which this learnability can emerge in the models. In the simplest case, a signal can survive the transmission bottleneck by becoming more generalisable. One of the ways in which this can happen is by becoming **compositional**; structured in such a way that the total of the meaning of the phrase is a function of the individual meanings of its constituent parts and the formal way in which it is arranged. As already discussed in §2.1, this feature is a key hallmark of natural language, and one which is largely responsible for the kind of open-ended generativity lacking in animal communication systems.

The way that generalisable utterances encourage their own survival lies in the fact that a compositional element can appear in multiple contexts, maximising its chances of being acquired (e.g., Kirby, 2000). When such a system emerges, it is not necessary to hear every possible utterance in the language, as the regular structure present in those utterances that were heard provides the learner with a method of reliably inferring the structure of those utterances that were not heard. For example, a child hearing 'red lorry', 'yellow lorry' and 'red car' could infer that there might be something called a 'yellow car' out there in the world on the basis of making a generalisation about the structural relationship between colour adjectives, vehicular nouns and their respective referents. They do not need to hear 'yellow car' paired with an actual referent to use or understand it.

A second way in which a signal can survive the bottleneck is by ensuring that it is used frequently enough to guarantee that the learner will hear it and need to acquire it. This insight is sufficient to explain another universal aspect of natural language structure: the presence of irregularity (Kirby, 2001). It is an interesting fact that in every language where there exist irregular forms, these forms tend to correlate with frequency of use in everyday speech. Thus for English, the top ten verbs are also all irregular (Francis & Kucera, 1982). In his model, Kirby (2001) manipulated the frequency with which certain meanings were sampled from the meaning-space, such that some were much more frequent, hence more likely to pass through the bottleneck, than others. What he found strongly mirrored the distributional patterns of irregulars in real languages: those meanings that were infrequent tended to be compositional, whereas those that were frequent were not. The message here is that we can learn idiosyncratic forms as long as they appear often enough in our input.

These models provide a proof of concept for the idea that language can adapt itself in response to the way in which it is culturally transmitted, and that some important structural features can emerge as a result of this dynamic. This perspective sees language itself as an evolving organism, capable of adapting to the environmental, social and cognitive pressures of its users. Whilst biological evolution has provided us with the necessary physiological pre-adaptations and much cognitive machinery for language (Hurford, 2003), it is not the sole adaptive mechanism at work. What gets acquired by one generation determines the data that future generations will use to construct their own language.

This has the advantage of taking away some of the explanatory burden from biological evolution – helping to account for some of the discrepancies involved, such as the incredible speed at which language is thought to have emerged[10] – whilst simultaneously incorporating our intuitive understanding of language as a cultural process. The key message to take home (Kirby, 2002a:27) is that: "(b)efore seeking a biological or functional explanation for a particular feature of language, or appealing to direct coding in an innate acquisition device, we should be aware of what we might be getting 'for free'"...via the mechanisms of cultural evolution.

In addition, one of the nice features of these models is that they do not commit us to any specific visions of how cultural evolution proceeds. So whether we view cultural evolution as a process whereby individual units of language get preferentially replicated (Blackmore, 1999; Croft, 2000; Aunger, 2002), or a process whereby the entire system is independently reconstructed anew at each generation (Sperber, 1996) is not important. Similarly, whether we choose to think of the learning biases as being language-specific or domain-general does not matter at this stage. The important thing is that iterated learning through generations can allow language to change, evolve and adapt culturally.

2.4 Cultural Evolution

Languages are undoubtedly culturally transmitted. The main aim of this thesis is to show exactly *how* this fact can actually explain why languages are structured the way that they are. We must start with the observation that language is not the only thing to be socially transmitted or evolve culturally in this way. Beliefs, skills, music, social attitudes, political systems, customs, architecture, religion, the rules of chess, fashion, mythologies, art and technology are also examples of things which arise and change over time as a result of cultural evolution. The diversity of behaviours,

[10] For views exploring this and other problems facing natural selection in explaining language evolution alone see Hurford *et al* (1998), and commentaries accompanying the publication of Pinker & Bloom (1990).

and in some cases, material artefacts that get classified as being 'cultural', or indeed, as forming 'culture' itself, is bewildering.

In a now famous survey, conducted in 1952, Kroeber & Kluckhohn examined the anthropological literature and found well over a hundred different definitions for culture alone. This has led to considerable divergence within the scientific community, with some studying culture-as-a-product (the customs, artefacts, behaviours and beliefs held by specific cultural groups), and others studying culture-as-a-process (the general mechanisms and adaptive dynamics that underlie this appearance of cultural products). Attempts have been made to bridge these two approaches. For instance, Richerson & Boyd (2005:5) define culture as follows:

> "Culture is information capable of affecting individual's behavior that they acquire from other members of their species through teaching, imitation, and other forms of social transmission."

By defining culture simply as information affecting behaviour, and jointly specifying the process by which it is acquired, they manage to bring together many of the different phenomena we would like to label as culture or cultural. This definition will also be adopted for the rest of the discussion here.

Obviously we know that language is not acquired through teaching or imitation. In fact, the previous section (§2.3) put forward the basic mechanism by which we see languages being culturally transmitted -- iterated learning. In this chapter I will be arguing that the fact that language is a relatively well-understood phenomenon makes it an ideal candidate for understanding processes of cultural transmission in general. However, that does not mean that researchers in language evolution should feel free to ignore the abundance of work undertaken by those investigating cultural evolution, thinking it only loosely relevant. On the contrary, even those studies that focus exclusively on teaching or imitation of non-linguistic behaviours in other species, can bring us closer to understanding what is essential for iterated learning, or language, or both. In particular, we will find towards the end of this chapter that the methods used by researchers to test predictions made by various cultural evolutionary theories will be of direct use to us here.

In some ways, linguists have been rather slow on the uptake. Research has been going on for a number of years investigating the relationship between language, culture and human cognition, but has largely gone unnoticed, perhaps because it has been deemed as fitting outside the bounds of proper linguistic enquiry. There are three main areas that have been explored: firstly, that complex language may have been a pre-requisite for complex culture; secondly, that evolving complex language may have actually enabled us to have more complex thoughts; and finally, the observation we have already noted concerning the fact that language itself arises as a result of a cultural process. The next section explores these three ideas in a little more detail.

2.4.1 The Relationship Between Language, Culture and Cognition

Language enables culture

Language is used to transmit cultural content in the form of ideas. Its capacity to do this has led evolutionary biologists John Maynard Smith & Eörs Szathmáry (1995; 2000) to conclude that complex societies with language represents the latest in a series of eight major evolutionary transitions in the history of life[11]. Each of these transitions typically involves some kind of aggregation (smaller entities coming together to form larger entities), division of labour, a change in the replication

[11] The first was the transition of replicating molecules to populations of molecules housed in cellular compartments; followed by the emergence of chromosomes from independent replicators; the transition from RNA to DNA; prokaryote cells to eukaryote cells; asexual cloning to sexual reproduction; single-celled organisms giving way to plants, animals and funghi; solitary individuals to colonies of individuals; before finally, the transition from primate society to human societies, which is heralded by the emergence of language (Maynard Smith & Szathmáry, 1999).

mechanism (after a transition, smaller entities that could once replicate independently can only do so as part of a larger whole), and the creation of new methods of information transference. The claim is that with the emergence of language, a whole new system of information transmission and replication appeared - one which, like DNA before it, supports unlimited heredity[12], and that this is what marked the transition from primate societies to human societies (Maynard Smith & Szathmáry, 2000). Language can therefore be seen as a powerful new evolutionary force in the world, giving rise to culture.

This latter point has been echoed by primatologists: one of the reasons why we have complex culture and our nearest primate cousins do not, is because only we have complex language (Boesch & Tomasello, 1998). Being able to encode information linguistically has been argued to make social learning more accurate, which is a necessary precondition for the emergence of cumulative cultural evolution and stable traditions (Sperber, 1996; Cavalli-Sforza, 2000; Atran, 2001). However, it is clear that not every culturally transmitted skill requires language. Shennan & Steele (1999) have argued that the manual skills required to generate stone tool technology could have been acquired simply through observation and without language. Similarly, Gil (2008) has questioned the argument that we needed complex grammar in order to acquire complex skills, such as building a boat and sailing it.

Language enables certain kinds of cognition

Perhaps then the value of language does not lie directly in what culturally acquired information it can transmit, but in the way it helps augment human cognition? One suggestion is that language gives rise to second order cognitive dynamics which help us make inferences about the world and ourselves – basically, the ability to evaluate our thoughts and plan our actions (Clark, 1998; 2006). Language from this

---

[12] Whereas DNA provides unlimited heredity to express biological information, language provides a system of unlimited heredity to express cultural information – ideas, beliefs, and skills etc.

perspective can be seen as a tool (part of our ’extended mind’), allowing us to freeze thoughts as objects which can then undergo scrutiny by the thinker, and more importantly, by other hearers. This has led some to argue that without language there are whole domains of abstract human concepts which could not exist, such as kinship relations, hypothetical situations, and ‘reasons’ for certain actions (Jackendoff, 1996). Similarly, the notion of a ’week’ (Pinker and Jackendoff, 2009), or even numbers (Hurford, 1987; Wiese, 2004) appear to rest upon language.

The claim is that simply having a mechanism by which we can transmit our thoughts and ideas to other people has fundamentally changed the way we think. Recent research has even shown that language can change the way our visual system works (Meteyard *et al.*, 2007; Winawer *et al.*, 2007; Lupyan, 2010), can influence our spatial reasoning abilities (Loewenstein & Gentner, 2005), and affects how we categorise novel objects (Vygotsky, 1962; Schyns *et al.*, 1998; Lupyan *et al.*, 2007). Understanding the extra-communicative roles that language may play in cognition may go on to help constrain theories of language evolution in useful ways (Lupyan, 2010).

Language is a product of the cultural process

This idea of language as a carrier, or vehicle for cultural information is not new. However, there is another sense in which we can see language and culture interacting; language conveys information about its own construction. That is to say, not only does language transmit culture, but it is itself also culturally transmitted (Brighton, Smith & Kirby, 2005; Kirby, Cornish & Smith, 2008). Children acquire language based on the previous output of the language learning of others, and this makes it a fairly unusual system (Zuidema, 2003). In a sense, it is equivalent to being able to infer the recipe and baking instructions of a cake, just by looking at it.

Interestingly enough, language is not the only system to have this property. It seems indicative of any traits that are acquired via a process of iterated learning. For instance, music and certain types of whalesong and birdsong also appear to cue their own construction in this way (Rendell & Whitehead, 2005; Feher *et al.,* 2009). Given that language is relatively well understood phenomenon, and humans are easy to run experiments on, this means that language can provide an excellent testbed for theories of cultural evolution more generally. It is to these theories that we now turn.

2.4.2 Theories of Cultural Evolution

Explicit parallels were drawn long ago between biological and cultural evolution (particularly, language evolution) by Darwin and his contemporaries:

> The formation of different languages and of distinct species … are curiously parallel … As Max Müller has well remarked : ' A struggle for life is constantly going on amongst the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their inherent virtue. ' (Darwin 1871:91)

Nevertheless, although we see the seeds of both disciplines emerging at the same time in history, the study of mechanisms of cultural evolution has lagged behind our understanding of the mechanisms of biological evolution by some magnitude (Mesoudi *et al.*, 2006b). There are many reasons for this, not least the fact that the field most closely associated with the study of culture -- anthropology -- has been strongly divided over whether something so rich and complex can be reduced to simple processes of cause and effect. Whilst many biologists would disagree with the implicit assumption that evolutionary theorising amounts to a reductionist explanation of a complex phenomenon (many biological processes are clearly more than the sum of their parts), others point out that some degree of reductive logic is no bad thing. To use the analogy developed by Dennett (1995), scientific theories that posit an over-abundance of cranes tend to explain whatever phenomenon they are attempting to explain, whereas theories resting on a single skyhook explain nothing.

Universal Darwinism

What are the parallels between biological and cultural evolution? For some, this is the wrong question to be asking. Instead, we should be concerning ourselves with understanding what general processes underly *all* forms of evolution. This quest to develop a general theory of evolution has been termed 'Universal Darwinism' (Dawkins, 1976; Dennett, 1995; Hull, 2001). In short, evolution is to be understood as involving three ingredients: variation, inheritance, and competition for survival. Any system where there is inherited variation of fitness is therefore an evolutionary one. Under this basic rubric, we can see that culture fulfills these criteria: we find variations between cultural traits, these cultural traits are passed on from person to person, and not all cultural traits can be expressed at once in an individual -- there is therefore competition between variants, not only within each individual, but between different population groups (Mesoudi *et al.*, 2006b).

Conceptual work linking Darwinism to culture has also been done by Mesoudi *et al.* (2004). Working directly from the text of Darwin's *Origins of Species* (1859), they suggest that a number of analogies can be found that go deeper than this. For instance, they point to shared features like convergent evolution, the presence of vestigial traits, the accumulation of modifications over time, the existence of adaptations *and* maladaptations, and similarities between the geographical distributions of species and the geographical distributions of certain cultural traits.

In spite of many similarities, there are differences to bear in mind. For instance, in some sense, any kind of cultural evolution is ultimately dependent on biology. This is true not only from the point of view of the mechanisms underlying cultural evolution requiring a biological explanation for their origins, but also from the point of view of cultural traits themselves. If a behaviour is maladaptive from the point of view of biology, that behaviour will not survive very long (Boyd & Richerson, 2005). For instance, Stone *et al.* (2007) discuss the case of the Albigenses, a religious sect that existed in Southern France in the 12th and 13th century. They believed that in order to attain pure spirituality one must abstain from marriage and reproduction entirely, and that since their material body was merely a cage for their soul, those Albigenses that wanted to attain perfection encouraged starvation and suicidal practices amongst themselves. Clearly it is easy to understand why this sect no longer exists today.

One particular issue that has received a lot of attention over the years is the units of selection debate: does culture consist of discrete units like memes (Dawkins, 1976; Blackmore, 1999), culturgens (Lumsden & Wilson, 1981), or linguemes (Croft, 2000) that get preferentially replicated in some way, or is cultural transmission more a process of complete reconstruction (Sperber, 1996; 2000; Atran, 2001)? If there are units of inheritance, at what level does selection operate? On the units of inheritance themselves, or on the individual possessing that trait, or even at the level of the cultural group that shared a trait? Although it is a divisive issue, which for some rules out any meaningful comparison between biological and cultural evolution (e.g. Bloch (2000) or Kupar (2000)), there is in fact little need for us to settle these issues immediately in order to develop testable theories. Whilst the general consensus seems to be that we should remain slightly cautious when making analogies with biological evolution, researchers have pointed out that Darwin himself was unaware of the precise mechanisms of inheritance when he developed his theory of natural selection (Aunger, 2000; Mesoudi *et al.*, 2006b).

If it turns out that there is no such cultural equivalent to a phenotype or a genotype, or that some cultural traits are directed towards a specific goal[13], then it is not a sign that culture is not 'evolutionary' (Mesoudi *et al.,* 2004). The whole idea behind Universal Darwinism is that biological evolution is only one type of evolutionary process. We should in fact predict that cultural evolution *will* have major differences to biological evolution. To summarise, all evolution requires is heritable variation of fitness. That is not to say that our understanding of biology is not relevant to our understanding of culture. As we shall see, even if we were to completely ignore any parallels between the mechanisms of biological and cultural evolution, it turns out that most theories of cultural evolution can be divided along the lines of how closely the process itself actually *interacts* with biological evolution. That is, any mechanisms of cultural evolution are enabled by biology at some level (Richerson & Boyd, 2005). Therefore we need a rudimentary understanding of how genetic transmission works even for this[14].

Evolutionary Psychology

Commentators classify evolutionary psychology (EP) as a theory of cultural evolution because it attempts to explain human behaviour as the result of processes of evolution (e.g Barrett *et al.*, 2002; Nettle, 2009). The EP approach links social and cultural behaviours tightly to biological underpinnings, with an emphasis on explaining variation in behaviours ("evoked culture") as the result of evolved

---

[13] This is a common criticism to be levelled at theories of cultural evolution. As humans are intentional beings, capable of planning their actions and innovating solutions to problems, the claim has been made by many that cultural evolution is the product of intentional design, and therefore fundamentally different to processes of biological evolution (e.g Hallpike, 1986; Pinker, 1997; Bryant, 2004). We will explore this issue in more detail in Chapter 3, but for now it suffices to say that the extent to which cultural evolution is directed or not is on a continuum - a matter of degree rather than strictly one or the other (Dennett & McKay, 2006).

[14] Unfortunately understanding the mechanisms of biological evolution requires a book in its own right, so rather than try to condense something that complex into a few paragraphs here, I instead direct the reader to Nettle (2009) for an accessible treatment of the subject.

psychological modules responding to different environmental inputs (Tooby & Cosmides, 1992). As such, it most closely parallels nativist explanations of language and language origins than accounts which give a more central role to cultural transmission.

This theory ties current behaviour to that which was adaptive in our past: the environment of evolutionary adaptation (EEA) - thought to correspond to some point in the Pleistocene (Tooby & Cosmides, 2000; Barrett et al., 2002). The argument is that certain behavioural traits which would have been adaptive for our ancestors (for instance, fear and avoidance of snakes) could, over time, have become genetically assimilated, as those people who possessed them were more likely to survive and reproduce than those who did not. Even though many of us live in an environment which is substantially different to the hunter-gatherer lifestyle of the EEA, these evolved behaviours continue to shape our current behaviour. For instance, we like sweet sugary foodstuffs now because those things would have helped us to survive in the EEA. Consequently, if we want to understand over-eating behaviour in current populations, we have to understand the role that such behaviour would have had in the past (Nesse & Williams, 1995).

Dual-Inheritance Models

Dual-Inheritance models (often alternately referred to as theories of Gene-Culture co-evolution) place their emphasis on the interactions between genes and culture. Unlike the EP approach, these models see culture as being currently adaptive, and transmitted rather than evoked (Nettle, 2009). As such, they fit more in line with the approach to language origins being advocated here. The idea behind these models is that culture and biology represent two distinct forms of inheritance, that can be functionally independent (Boyd & Richerson, 1985). In spite of their relative independence, they can also interact with one another in interesting ways. Culture can affect genes directly, as for example, in the link between dairy farming and lactose tolerance (Durham, 1991). Conversely, genes can affect culture directly too., as any cultural trait that is deleterious to the reproduction of the organism will be wiped out (remember the example of the Albigenses people in the previous section). However, beyond this, there is a whole raft of possible interactions between the two.

Richerson & Boyd (2005) in particular emphasise the importance of population effects. Cultural traits can spread because they affect an organism's biological fitness, but they can also increase because they affect an individual's cultural fitness. Ultimately for Richerson & Boyd, all tributaries lead to the sea: cultural fitness can act as a proxy for biological fitness as much as the elaborate Peacock's tail (Zahavi & Zahavi, 1997). Nevertheless, within this cluster of theories there is greater emphasis placed on culture being free to evolve for culture's own sake.

Niche Construction

Niche construction theories also deserve a brief mention. Sometimes referred to as 'trial-inheritance theories', these can be seen as a kind of extension to the dual-inheritance theories described earlier. As well as biological and cultural inheritance, proponents argue that there is also a third system of ecological inheritance (Odling-Smee *et al.*, 2003). Not only do organisms adapt *to* their environments, but they also adapt their environments (Stone *et al*, 2007). Classic examples of the basic principle include beaver dams or bird nests, which effectively change the environment in which an organism must survive. These changes tend to last longer than the organism itself, are sometimes literally inherited by their offspring, and may even impact upon different species altogether. Long-term changes in the ecological niche inhabited by an organism in turn effect the selection pressures operating on that organism - for instance, the presence of a beaver dam creates pressures for beavers with certain morphological features, like stronger teeth and flat tails (Laland & Odling-Smee, 2000). Our theory of language evolution fits in quite well with this model, as the actual language being transmitted is itself an environment of sorts, that goes on to affect future generations.

As Bullock & Noble (2000:150) note in a discussion of the relevance of Kirby & Hurford's (1997) model of language evolution to niche construction:

> "New-born organisms must learn a grammar from a set of utterances provided by the parental generation. Thus the ecological legacy is not the physical environment but the linguistic one: a new organism is born into a world of speakers."

There is definitely something captured by the theory of niche construction that is shared with those theories of language evolution that stress the importance of iterated learning; namely the great emphasis that both theories place on interaction and selective pressures arising at many different levels. Work has already begun on making those parallels clearer (e.g. Odling-Smee & Laland, 2009).

2.4.3 Modes of Cultural Transmission

Working from mathematical models of biological evolution, Cavalli-Sforza & Feldman (1981) identified three different directions that cultural transmission could proceed in: vertical, oblique or horizontal. Of all these, **vertical transmission** shares the closest parallels with biological evolution, as it relates to the way in which cultural information gets passed down from parents to their children. Similarly, **oblique transmission** also refers to information passed down from generation to generation, but rather than specifying a parental relationship, this refers to any interactions between adults and children, or where information passes from someone with experience, to someone with less experience[15]. Finally, **horizontal transmission** relates to information being passed between members of the same generation, and/or level of expertise.

[15] Due to the fact that vertical and oblique both refer to inter-generational transfer of information, many authors use the term vertical to refer to them both. In general I will follow this convention as well.

Various claims have been made about the different properties that each of type of transmission has. For instance, vertical transmission has mostly been associated with conservation and stability of traits (Laland *et al.,* 1993), with prime examples of this being language and hygiene practices which tend to correlate strongly with those held by the parental generation (Stone, *et al.,* 2007). Empirical evidence for this also comes from studies of Iranian rug-making, which reveal how mother-daughter transmission results in extremely stable designs (Tehrani & Collard, 2002), and studies of Stanford grad-students showing certain cultural traits which do not tend to change over time, like voting preference, are acquired vertically (Richerson & Boyd, 2005). In contrast, horizontal transmission has been associated with the generation of innovations and variation, and the rapid spread of cultural information (Stone *et al.,* 2007).

Even within these broad types of transmission identified by Cavalli-Sforza & Feldman (1981), there is additional recognition of some sub-types. For instance, within horizontal transmission Stone *et al.* (2007) differentiate *one-to-one* (the 'standard' form of horizontal transmission as conceived by many researchers), *many-to-one* (a more powerful form of transmission where several people transmit the same information to just one individual -- commonly described by psychologists as 'peer pressure'), and *one-to-many* (where a particularly prestigious individual -- for instance, a political leader or celebrity -- influences the spread of information on a large-scale). Each of these sub-types also have different effects, with many-to-one transmission being particularly hard to resist, and one-to-many being associated with very rapid cultural change.

Despite a wide-spread belief that the majority of cultural transmission is horizontal, there is a surprising amount of anthropological evidence for vertical transmission. Researchers have used the fact that the different modes of transmission result in different distributional patterns in order to identify which direction is the more dominant. In spite of the fact that cultural evolution definitely allows a greater scope for horizontal transmission than genetic evolution[16], vertical transmission is often still the dominant mode in cultural evolution (Guglielmino *et al.,* 1995). This was demonstrated by a study which analysed some 47 different cultural traits in 277 African communities. Guglielmino *et al.* (1995) reasoned that if horizontal transmission is more dominant, we would expect that cultures would tend to share traits with those communities geographically adjacent to themselves, but if vertical transmission is stronger, we would expect groups to conserve the traits of the cultures they descend from. It was statistically shown that the majority of traits showed evidence of descent over generations[17] – especially amongst those traits most closely connected to reproductive success. This study was later supported using a worldwide sample (Holden & Mace, 1999). Collectively results like this have been taken to show that:

> "even under the influence of close geographical neighbours, cultures can remain stable and coherent units...cultural evolution is not a free-for-all in which all traits become equally available for adoption each generation." (Pagel & Mace, 2004:277).

In other words, the concepts of distinct lineages and restrictions on descent are strongly operative within cultural evolution. This finding also goes some way towards addressing the concerns described earlier about whether cultural evolution can ever be fully understood if there are no clearly identified, discrete 'units' of selection.

In the end, what all of these 'directional' accounts boil down to is that differences in population structures can substantially affect the dynamics of transmission, even if the underlying mechanisms remain the same. Mesoudi (2007) and Mesoudi &

---

[16] It turns out that there are horizontal transmission type processes operating within biology, particularly with regards to the immune system (Jablonka & Lamb, 2005).

[17] The main exception to this trend were traits related to sexual division of labour, which did seem to pattern more with horizontal transmission.

Whiten (2008) agree. In a review of different cultural evolution experiments (see $2.4.4) they identify three basic types of population structure: linear diffusion, closed groups and replacement. In the broad terms we have been using so far, these correspond to vertical, horizontal and vertical and horizontal combined. The advantage of adopting this terminology and thinking about the situation in terms of population structure rather than direction of transmission however is that we can avoid falling into the trap of thinking of vertical transmission as being more Darwinian and horizontal as being more Lamarckian. This is important, as we have no reason a priori to assume that the mechanisms underlying transmission differ when one is engaged in cultural exchange with someone from your parent's generation, and someone from your own. The next section will look at some actual examples of cultural evolution experiments that implement these various methods in more detail.

2.4.4 (Non-Linguistic) Cultural Evolution Experiments

Experiments on cultural evolution have mostly focused on identifying the precise social learning mechanisms that enable the development and maintenance of cultural traditions, in both humans and non-human animals. Until fairly recently, the idea of animals having any kind of cultural tradition would have seemed very strange. This has changed however, as studies of animals in the wild have revealed that not only do many species have a rich cultural life (e.g. McGrew, 1992; Boesch, 1996; Whiten *et al.,* 1999; Watanabe, 2001), but that the ability to transmit simple traditions between conspecifics is not unique to primates. It has in fact independently evolved several times in other species (Laland & Williams, 1997; Rendell & Whitehead, 2005; Janik & Slater, 2003; Fitch, 2010). This tells us that social learning is adaptive, and supports models and experiments of the process which show under which circumstances social learning offers the clearest advantages over individual learning (Boyd & Richerson, 1995; Kameda & Nakanishi, 2002).

One of the main focuses of interest lie in what types of social learning mechanisms support **cumulative cultural evolution**. One of the interesting differences between the types of cultural traditions that non-human animals have, and the types of cultural traditions that humans have, is that ours are said to involve the accumulation of innovations. Animal cultures, in contrast, are perhaps more fittingly described in the way that Boyd and Richerson view cultural inheritance "as a shortcut to individual learning" (Boyd & Richerson, 1985:14). Animals learn things socially that they could have discovered by themselves via trial and error individual learning. Humans on the other hand socially learn things that are too complex for them to have discovered independently (Boyd & Richerson, 1995).

Tomasello (1999) attributes this difference in complexity to the 'ratchet effect': a combination not only of creative invention, but of social transmission that has a high degree of fidelity to prevent backward slippage and allow new innovations to be faithfully preserved and accumulate complexity over time. Consequently, focus on the mechanisms of cultural evolution has paid a great deal of attention to those which support high fidelity transmission, such as imitation and teaching, whereas those like stimulus enhancement, emulation and ontogenetic ritualisation have been argued to be insufficient for cumulative cultural evolution (for discussion of these terms in this context, see Tomasello *et al.,* 1993).

Recently research has been less focused on identifying the precise mechanisms that support social learning in different species, and more on how any cultural trait, simple or cumulatively complex, gets transmitted through populations of individual learners. Whilst it has occasionally been possible to get close enough to observe populations of animals interacting socially in the wild (Biro *et al.,* 2003), the greater experimental control of laboratory studies is often preferred (Whiten, 2005). Broadly speaking, there are three experimental methods that have been used by social and comparative psychologists to study cultural transmission (Mesoudi, 2007; Mesoudi & Whiten, 2008). Figure 2.7 shows these different methods. The rest of this section will explore some examples of each type of experimental method.

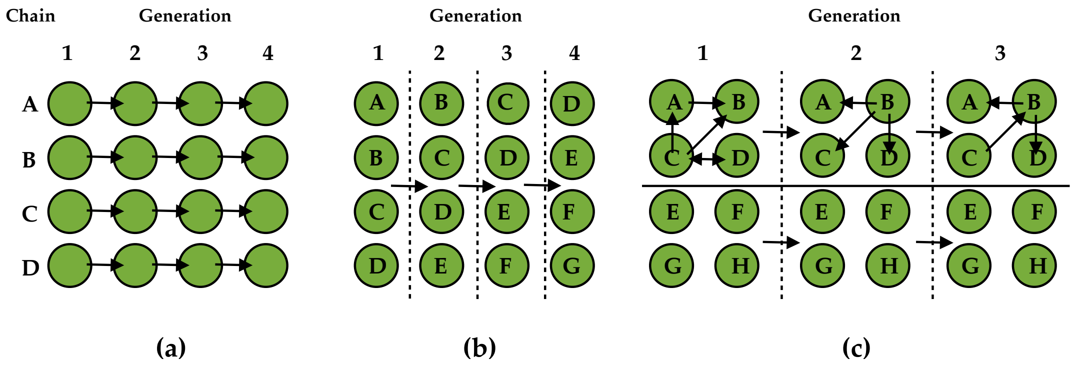


Figure 2.7: Three different experimental designs for cultural transmission experiments. Participants are represented by circles, and arrows indicate the direction of transmission either between generations (arrows cross vertical dotted lines), or between individual participants. In (a) we see the design for a typical transmission chain study, where information is passed along parallel chains (indicated by letter) of participants. In (b) we see the design of a typical replacement study, where four participants (A-D) interact together in some learning task. One participant gets replaced by a new learner at each generation. In (c) we see the design of a typical closed-group study. There are two conditions. In the upper, we see the social condition, where four participant (A-D) repeatedly engage in a learning task together. In the lower section we see an individual learning control condition, where participants engage in the same learning task, but are not allowed to interact with one another. Reproduced from Mesoudi (2007).

Linear transmission

In the **linear transmission chain** method, participants are organised into different chains, and information is passed along like in the game 'Chinese Whispers' or 'Broken Telephone', with each learner learning from the previous. This corresponds to the broad definition of vertical transmission described in §2.4.3. This type of experiment has a long history of use in human social psychology, most prominently being used to explore how people's recall of narrative descriptions change over time depending on their cultural expectations or pre-existing knowledge (Bartlett, 1932; Allport & Postman, 1947; Bangerter, 2000; Mesoudi & Whiten, 2004) or how prior cognitive processing biases impact upon information being transmitted (Kalish *et al.*, 2007; Griffiths *et al.*, 2008). Often referred to as diffusion chains, this technique has also been used to explore how humans and non-humans can sustain different foraging traditions.

In one type of experiment, learners are given the task of opening a puzzle box containing a food reward. This is sometimes referred to in the literature as an artificial fruit (Whiten *et al,.* 1996; Custance *et al.,* 2001; Caldwell & Whiten, 2004). Typically, there will be multiple ways of manipulating the puzzle box in order to access the food, but learners will only be shown one method. These methods can be thought of as learned cultural variants. Traditionally, these studies have been used to isolate whether the species involved is capable of acquiring a cultural variant via observational learning, or whether the behaviour is learnt via individual learning techniques. However, by using a slight twist of the linear transmission method, known as open diffusion, researchers can use the artificial fruit task to examine how culturally acquired behaviours can actually be passed on through separate groups of primates (including human children).

For instance, Whiten (2005) took three different groups of captive chimpanzees, and exposed them to a 'pan-pipes' device containing grapes. In two of the groups, a high ranking female was taught a technique for opening the device by a human demonstrator. The first technique involved using a stick to lift a catch, whilst the second involved using a stick to poke a release mechanism. In the third control group, neither technique was demonstrated. Once the chosen chimpanzee had acquired one of the two variants (lift or poke), the pan-pipes and the chimpanzee were returned to the group. Over the next few weeks, researchers noted the interactions between different chimps, and tracked how the modelled behaviour spread through the group.

One of the perhaps surprising results of studies such as these is that not only do chimpanzees and other primates show a strong bias towards conformity, preferring to adopt whichever technique is used by the group as a whole even if other techniques are independently discovered during experimentation (Whiten, 2005; Dindo *et al.,* 2009), but that (chimpanzees at least) also prefer to copy the most prestigious model (Horner *et al.,* 2010). This mirrors similar findings in human adults concerning conformity and prestige biases (see Richerson & Boyd (2005) for discussion of these biases in general), suggesting that this could be a capacity shared with a common ancestor.

The more standard linear diffusion chain experiments have also been conducted on children using the artificial fruit method (Horner *et al.*, 2006; Flynn & Whiten, 2008). These studies not only found the same conformity bias and faithful transmission of technique as in previous studies, but that there are both developmental and gender differences as well: older male children are better at faithfully imitating complicated behaviours than younger females. Interestingly, whilst studies involving dyadic transmission of behaviours from an adult to 3 year-olds have shown that children tend to over-imitate (i.e. copy even obviously redundant actions when trying to open a puzzle box or follow a recipe) (Horner & Whiten, 2005; Gergely & Csibra, 2006), when interaction continues beyond the dyad and along a transmission chain of other children, this redundant information is rapidly parsed out (Flynn, 2008). This contrast between the behaviour of individuals engaged in a 'one-shot' learning task, versus the behaviour of multiple individuals engages in the same learning task over multiple generations, is a persistent finding in cultural transmission studies conducted in the laboratory, and something we will come back to in §3.2.3.

Replacement

In a somewhat different experimental set up involving just human participants, researchers have attempted to create 'microsocieties' in the laboratory[18]. This type of experiment typically involves the kind of **replacement** method illustrated in Fig. 2.7. In these studies, groups of participants interact with one another whilst performing some task. After a while, each group member is replaced one by one, with each replacement representing a new 'generation'. Unlike linear transmission then, there is continuity of participants between generations. The combination of interaction,

---

[18] Not every microsociety experiment involves the replacement method. The next subsection describes microsociety experiments which use the closed-group method.

and generational turnover means that this method can be seen as a combination of horizontal and vertical transmission.

The replacement method also has a fairly long history within social psychology, most often being used to investigate group conformity and how long it takes for negotiated or experimentally induced social norms to break down (Gerrard *et al.*, 1956; Jacobs & Campbell, 1961). More recently, it has been used to show how interacting groups and chains of participants can develop optimal behaviours over time (Baum *et al.,* 2004; Caldwell & Millen, 2008), or under what conditions participants rely more heavily on social rather than individual learning (Caldwell & Millen, 2010). In the set of studies conducted by Caldwell and Millen, groups of participants are given the task of either making a paper aeroplane that will fly the furthest, or building the tallest tower out of spaghetti and modelling clay. We will briefly look at some examples of these.

In the first spaghetti towers study (Caldwell & Millen, 2008) seemingly arbitrary designs were found to emerge over time in each transmission chain. At any one time during the study there are two participants building towers, and two participants observing them. When 'builders' have finished, they are replaced by new 'observers', and the old observers become builders. This process continues along the transmission chain for a number of generations. The similarity between the resultant tower designs created at each generation was then rated by independent observers. This similarity was found to be greater within-chains, than across-chains. In other words, tower designs were being passed on by individuals within the chain. In addition to this, towers were seen to increase in height cumulatively, as learners selectively retained elements of good tower design from previous participants in the chain.

In the second study (Caldwell & Millen, 2010), uncertainty is introduced by the additional requirement that, after construction, the tower must be placed next to a desk-fan for five minutes. As towers have a tendency to collapse not long after being built, this creates a situation where participants are less certain as individuals about what constitutes a 'good' design. In this condition it was found that participants relied much more heavily on the design of the previous builders than before. Additionally, and unlike what was found in the previous study, there was no significant increase in the height of the towers over time. One possible explanation for this is that the greater reliance on social learning is in some way inhibiting individual innovation, which is necessary for cumulative cultural evolution.

Closed-Group

This method explores cultural transmission between learners where there is no generational turnover at all. It therefore most closely corresponds to what has been termed horizontal transmission. Again, these experiments are often referred to as microsociety studies. For instance, McElreath *et al.* (2005) and Mesoudi and O'Brien (2008) both investigate how individual learners modify their strategies based on observing how other individuals react in similar environments. These studies are microsocietal in that participants are making choices about how to perform some function as a group: in McElreath *et al.,* (2005) participants are given the role of farmers trying to maximise crop yields, whereas in Mesoudi & O'Brien (2008), participants are designing the optimal arrowheads for hunting.

In both of these studies, participants were given the chance to examine the behaviour of other members of the microsociety and modify their own behaviour in response. For instance, in McElreath *et al.* (2005) participants could view what crops other farmers had chosen to plant. Against the predictions of models, the study found that a large number of participants did not take advantage of cultural learning, even when it would have resulted in a greater crop-yield. Of those that did copy, they only chose to conform to the behaviour of others when the environment changed and they were no longer getting an optimal pay-off. This indicates not only that there is a substantial amount of individual variation in the willingness to conform, but also that models of when people are likely to switch social learning strategies are not always accurate. Human participants do not always behave optimally.

2.5 Summary

This chapter began by looking at what features make language an interesting phenomenon to understand. It argued that in order to understand how languages are acquired with such reliability and ease, and why languages of the world all share similar structural properties, we need to understand how languages evolved. Two contrasting accounts for language evolution were presented. In the first, universal properties of linguistic structure were seen as the direct consequence of genetically determined language-specific learning biases. In the second, universal structural properties were held to be emergent, arising from the interactions between biological evolution, individual learning, and cultural transmission.

The fact that languages are culturally transmitted has been argued to at least partially account for why they exhibit the structural properties that they do. Language is the result of a process of iterated learning. Iterated learning has been extensively studied using computational models. I discussed the two main findings to have emerged from these studies: that languages adapt to be easier to learn over time, and that they also adapt to convey structured meanings by becoming structured themselves. We then moved on to explore cultural evolution more generally.

We began by exploring three different relationships that language shares with culture and cognition, pointing out that not only is language the conveyer of cultural content, but is itself the product of cultural processes. Language may also have enabled higher-order cognitive functions by virtue of providing a mechanism to share thoughts with others. Next we briefly examined some of the parallels and divergences between biological and cultural evolution, noting that whilst there is much disagreement about how cultural evolution is actually instantiated (for instance, whether it has discrete units of inheritance what the unit of selection is), it is not necessary to understand the precise mechanisms of inheritance in order to get an understanding of how evolution might proceed.

At this point, three different theories of cultural evolution were introduced: evolutionary psychology, dual-inheritance and niche construction. Each of these theories differs in the extent to which biology can be thought to dominate, and the extent to which human behaviour can be thought of as being currently adaptive. The evolutionary psychology approach, with its emphasis on evolved cognitive modules, was argued to be more compatible with explanations of language origins that make a direct appeal to biology. Both dual-inheritance and niche construction theories on the other hand emphasise the role of interactions between separate forms of inheritance: biological, cultural, and in the case of niche construction, also ecological. Language is a particularly good example of niche construction, as it is itself a kind of inherited environment that lasts a good deal longer than many of its speakers.

The next topic to come under scrutiny was the different modes of cultural transmission. Traditional definitions have focused on making distinctions between vertical (inter-generational) and horizontal (peer-to-peer) transmission. However, because of the long-standing assumption that vertical transmission is 'like biology' and horizontal is 'like culture', it was suggested that a better way of thinking about modes of transmission was in terms of the structure of populations rather than purely by direction. This is in part due to the fact that cultural transmission is not particularly dominated by horizontal exchange, and also because experimental psychologists studying cultural evolution in the laboratory have been using different terminology for a number of years.

Finally, this chapter reviewed some of the literature on a sample of those empirical investigations of cultural evolution in the laboratory. These studies have revealed many interesting continuities and differences between species, and also shown under what conditions we can expect to see social learning strategies favoured by participants, and how cumulative cultural evolution can be investigated in our own species. However, none of them have examined the topic of linguistic transmission. This is the focus of the next chapter.

Chapter Three

Empirically Investigating Language Evolution

It seems then that there are a number of ways in which the cultural transmission of information has been explored in humans, as well as non-humans[1]. However, none of the experiments we have examined so far have made language itself the empirical target. This chapter will introduce research that does just that. In reviewing the existing literature, it aims to motivate a new experimental methodology for studying language evolution in the laboratory. It first examines why laboratory based experiments have only really been developed over the past few years, and describes some of the problems with investigating the origins of language empirically. It then moves on to look at some of the current approaches that have been successful, including computational studies, observational studies of natural language emergence, artificial language learning studies, and finally, experiments involving the emergence of artificial systems of human communication.

The question of intentional design in language will then be approached. I will argue that there is a potential issue with the way in which current laboratory experiments investigate the emergence of novel communication systems, which allows for the participants to intentionally design a communicative system. This is not a good model of language evolution for several reasons. I will then outline the general

[1] Parts of sections 3.2.4, 3.3 and 3.4.3 of this chapter appear in Cornish (2010).

methodology for an experimental framework that specifically rules out the possibility of learners intentionally creating systems designed for communication. Particular attention will be paid to the way in which results from this framework can be analysed, ahead of the experimental results which will appear in Chapters 4-6.

3.1 What took you so long?

It may come as a surprise to researchers in other fields that the study of language origins has only recently started to collect data from laboratory experiments. Given the close relationship evolutionary linguistics shares with fields such as psychology, computer science, biology and developmental linguistics -- all fields associated with a high degree of empirical investigation -- it is more surprising still. However, evolutionary linguistics also has close ties with disciplines such as philosophy and cognitive science which, perhaps unfairly, have traditionally been associated with integrating empirical results from other fields in order to fashion out new theories, rather than generating empirical results on their own.

It is possible that, to some extent, we are still seeing the after-effects of history. Both prior to and immediately after the publication of *The Origin of Species* (Darwin, 1859) there was much interest in the evolutionary study of language. Due to the wildly speculative nature of the theories that emerged during this time period, unconstrained as they were by any firm knowledge of language acquisition, genetics and neurological processing which might have limited theorising to the realms of the more plausible, in 1866 the *Société Linguistique de Paris* enacted their famous ban on the study of origins and evolution of language (Christiansen & Kirby, 2003). This put a stop on this area of research for the next hundred years or so, and perhaps still affects the way evolutionary linguistics is viewed today.

Whatever the cause, the sentiment that evolutionary linguists cannot employ traditional empirical methods has led at least one notable practitioner to recently conclude in a major linguistics journal that:

> "To enter [the field of language evolution] costs little: you can't do experiments, so no expensive equipment is required...It's still a pencil-and-paper field" (Bickerton, 2007: 524).

Lee *et al.* (2009: 32) have also made similar remarks concerning the impossibility of studying language evolution in the lab: "[I]t is not possible to use real human beings in experiments to see whether linguistic structures can emerge through simple interactions." This idea that we cannot investigate language evolution using human participants is false, as the work that follows will show. However, the assumption still lingers, especially amongst researchers working just outside the field. Perhaps what is really at the heart of the problem is that language evolution presents a unique problem to science. How do we study the emergence of something so complex and rare that it has only happened once in the history of the world? How do we even begin to approach a problem that happened so long ago?

The Difficulty of Studying Language Evolution

Linguists are well used to viewing language as a formal, idealised, and rule-governed system. However, when we consider language as a complex adaptive system (CAS) things start to get decidedly non-linear. This is because in CASs (such as language) the total rarely equals the sum of its parts. Simple local interactions often give rise to complex emergent behaviour (Johnson, 2001). Furthermore, evolution is necessarily a historical process, which means that there may always be some element of randomness about it (de Boer, 2005). Historical accidents appearing early on in time can remain 'frozen' and constrain future development in fundamental ways, and even slight differences in the initial conditions can result in massively different outcomes in the final product (Gell-Mann, 1994).

These are just some of the difficulties that lie with any attempt to uncover the truth about language evolution. The sheer complexity of the phenomenon aside, efforts are also hampered because the object of study is not even visible to us – there is no way of going back in time or recreating the exact conditions that led to the emergence of language in our hominid ancestors (Christiansen & Kirby, 2003). Even if we could somehow go back in time with a team of researchers, we would have very little idea of what year we should return to. Even identifying roughly *when* the capacity for language evolved has proven a difficult challenge, let alone understanding *how* it evolved. We know that language must have been in place before anatomically modern humans left Africa, some 50,000 years ago, but tracing the capacity for language beyond this has proven problematic (Mellars, 2006).

Tracing Language(s) Through Time

One approach has been to examine the fossil record for clues to when language might have emerged. Unfortunately archaeological data cannot give us any direct clues, as language, being non-physical, leaves very little trace (Hauser & Fitch, 2003). With that caveat in mind, some researchers have looked for clues in the skull structures of early hominids. One of the notable features of *Homo sapiens* is the brain size to body size ratio. An oft-quoted figure is that our brains are three times larger than we should expect for an ape of our size (Fitch, 2010). An increase in the size of our brains relative to our closest neighbours has long been associated with an increase in cognitive abilities, although this has been called into question (e.g. Macphail, 1982; Deacon, 1997). The message seems to be that bigger brains might have more processing power, but this might not correlate to more sophisticated behaviour or, more importantly, linguistic behaviour.

If we cannot learn anything about language evolution from the structure of the brain, what about other structures that are vital for language, such as the vocal tract? Although fragile and not particularly well-preserved over time, the shape and positioning of the hyoid bone in certain specimens of *Homo neanderthalensis* has led some researchers to conclude that Neanderthals probably had the same range of speech sounds as modern humans (Arensburg *et al.,* 1989; Boë *et al.,* 2002). However, this is contested (Lieberman, 2007), and in any case having the ability to make speech sounds is not the same as having the ability for language. After all, we know that chimpanzees are capable of making some of the gestures of sign language, but even after extensive training they still cannot fully acquire it (Gardner & Gardner, 1969). Work investigating the vocal production in other species, such as dogs and deer, has also revealed that most mammals have a more dynamic vocal tract than previously thought which allows them to radically reconfigure their vocal anatomy when vocalising (Fitch, 2000). This once again urges us to be cautious in attempting to form conclusions based on fossilised evidence.

A different line of enquiry however has been to look at the archaeological record in terms of material culture. In other words, can we learn anything about language evolution by looking at the kinds of artefacts our ancestral hominids were making, or any evidence of their behaviours that they might have left behind? Judging from the discovery of accumulations of animal bones, and the analysis of stone tools, it seems reasonable to suggest that by 2 million years ago hominids were sharing food with one another and being sociable (Isaac, 1978; Plummer, 2004). We also know from fossilised footprints that hominids were bipedal at least 3.6 million years ago (Leakey & Hay, 1979). There is also archaeological evidence for what has been termed an 'explosion' in material culture 40-60 thousand years ago, heralding not only an increase in the number and designs of functional tools, but also the emergence of symbolic artefacts, such as art and decorative pieces (Deacon, 1997; Lewin, 2005).

From all this indirect evidence, Barrett *et al.* (2002) present three possible scenarios of when language may have evolved: (a) in early *Homo erectus*, around 1.5 to 2 million years ago (e.g. Deacon, 1997); (b) when *Homo sapiens* first appears around 500 thousand years ago (e.g Falk, 1980; Aiello & Dunbar, 1993; Worden, 1998); or (c) around the time of this material culture explosion, 40-60 thousand years ago (White, 1982; Noble & Davidson, 1996).

Rather than looking for historical evidence of biological hallmarks of language emergence, can we learn anything by examining how individual languages have formed over time? The process of language reconstruction has a long and distinguished history within linguistics, although attempts to reconstruct earlier forms of language based on similarities between extant languages can only go back so far (Fox, 1995)[2]. Related to this, other approaches have looked at genetic data and the distribution of current languages, in conjunction with what is known about human migrations and population expansions throughout history (e.g. Cavalli-Sforza, 2000)[3]. Although the goal of this work is to understand more about human history and evolution in general, the outcome of such studies does help to constrain theorising about the evolution of language as well.

The main issue with both linguistic reconstruction and attempts to study human evolution over these shorter time-frames is that we run the risk of investigating language *change*, rather than language *evolution.* The difference between the two is subtle, but important. Whereas language change involves systems moving through the space of possible linguistic states, language evolution involves systems moving *between* spaces of possible linguistic states themselves. That is, it involves the transition from a state of no language to a state of language, rather than a transition from a state of language to a slightly different state of language (see Brighton (2003) for more discussion of this distinction). Whilst learning more about the ways in

[2] Note that in spite of the note of caution sounded by researchers like Fox (1995) attempts have been made to reconstruct languages as far back as 'Proto-World', the hypothesised language from which all modern languages are descended from (e.g Bengtson & Ruhlen, 1994; Ruhlen, 1994).

[3] This kind of research is different to the work linking recent genetic changes in human DNA to specific features of language (Dediu & Ladd, 2007) in that it attempts to use genes of modern populations as a historical record of the human species. It is thus much more like linguistic reconstruction in nature, and focuses on language change, rather than language evolution.

which current languages change over time is relevant to our study, we must be careful to keep in mind that our original goal is to explain the emergence of 'language', not specific languages.

Potential for Progress

One positive sign of progress comes from the comparative studies we explored in Chapter 2. Not only have the comparative studies of animal communication been useful for helping us identify which features of language are uniquely human, but we can also learn a lot from the degree to which our biological cognitive foundations for language are shared with other animals. For instance, determining whether the trait is homologous (i.e. related by descent) or analogous (i.e. arising independently in a separate lineage) can tell us whether that trait is present for functional or historical reasons (Fitch, 2010). In many cases, this alone is sufficient to tell us something about the evolutionary pressures driving selection, most obviously, whether or not that trait is an adaptive response to pressures arising from similar environments.

Another way in which comparative studies can inform research into language evolution is by telling us something about how culture evolves, or social learning behaviours in general (e.g. Boesch & Tomasello, 1998; Caldwell & Whiten, 2006; Whiten & Mesoudi, 2008). Although data from animal studies is definitely relevant to addressing questions about language evolution (and an interesting topic of study in its own right), it is still only indirectly related to the phenomenon we wish to understand. We cannot always assume there is a straightforward relationship between what we learn about animal communication and social learning mechanisms, and our own capabilities. However, one thing is certain: if we can get empirical data on how processes of cultural evolution work in non-humans, we should also be able to get empirical data on how processes of cultural evolution give rise to language in humans. Indeed, we have also seen several examples of this type of experiment in both animals and humans (§2.4.4). Although these did not involve language, they still set a useful precedent for studying some of the mechanisms of cultural transmission that support linguistic transmission.

Perhaps the most important thing to bear in mind when thinking about ways in which we might progress our understanding is that we have the evolutionary end-points of the process (i.e. modern languages) to hand. Even as you read this, languages are evolving[4] -- although it is rare, we do have some limited access to natural cases of language emergence that are ongoing today (e.g. Nicaraguan Sign Language, creolisation). With greater constraints provided by our knowledge of neurology, language acquisition, language disorders, plus insights that can be gained from the analysis of computational simulations and formal modelling techniques, progress *is* being made in the field of language evolution. The next section explores some of these avenues.

3.2 Methods for Studying the Cultural Evolution of Language

Recall that the main aim of this thesis is to explore how language evolves as a result of being culturally acquired via iterated learning. Therefore our focus in this section will be on methods for understanding cultural evolution only.

3.2.1 Computational and Mathematical Studies of Language Emergence

As discussed in the previous chapter, over the last few decades the use of computational simulations and mathematical models to explore language evolution has rapidly increased. The advantage of this methodology stems from the fact that models allow us to check and refine our theories very rapidly. One of the key issues with studying CASs is that our intuitions do not always naturally match up with

[4] If we are being picky, with the exception of newly emergent systems, we should really say that most languages are merely *changing*. See Brighton (2003) for a discussion on the differences between language change and evolution.

reality (Hashimoto, 2002). Making our theoretical assumptions explicit in a formal model of the process allows us to rigourously test whether our predictions do in fact follow from our hypotheses. Of the various models out there, the most relevant to the current work are the ILMs discussed earlier (see §2.3), which focus on explaining the emergence of compositional structure in language in terms of cultural transmission (e.g. Kirby & Hurford, 2002; Smith *et al.*, 2003; Brighton *et al.*, 2005). However, there have also been a range of computer simulations which explore the emergence of innate signalling systems as a result of purely biological evolution[5]. In particular, these studies have focused on determining under which ecological conditions we can expect to see evolution by natural selection resulting in the emergence of simple communication systems (MacLennan & Burghardt, 1994; DiPaolo, 1997; Cangelosi & Parisi, 1998; Noble, 1999), or on understanding the origins of the communication channel itself (Quinn, 2001). There have also been models conducted which explore how both cultural learning and biological evolution can interact together (Hinton & Nowlan, 1987; Kirby & Hurford, 1997; Watanabe *et al.*, 2008) -- therefore focusing on all three elements of complex adaptive system described in §2.2. This can give us valuable insight as to how iterated learning may fit into the bigger picture of language evolution as a whole.

Additionally, the problem of language emergence has also been investigated mathematically (e.g. Niyogi & Berwick, 1997; Nowak *et al.*, 2002; Griffiths & Kalish, 2005, 2007; Kirby *et al.*, 2007; Griffiths *et al.*, 2008; Ferdinand & Zuidema, 2009). Many of these more recent studies have focused explicitly on separating the respective contributions of the process of transmission and the pre-existing learning biases held by the agents. This has been achieved by modelling agents as Bayesian learners, who form hypotheses about the data they have seen based not only on the likelihood of that hypothesis actually having produced that data but *also* the prior probability of that hypothesis being entertained by the agent without having seen

[5] See Oliphant (1997) or Kirby (2002) for a more in-depth review of these models.

any data (Griffiths & Kalish, 2005; Kirby *et al.*, 2007; Smith & Kirby, 2008; Ferdinand & Zuidema, 2009; see also §5.3 for discussion of these models).

Despite the breadth and depth of this research method, it is not immune to criticism. Although it is always possible to find specific faults with individual models, there is one charge that has been made toward computational models in general: that they over-simplify their subject matter. In some sense this is what makes the models desirable – we use models when we want to grasp the underlying dynamics of complex phenomena, and to do this, we must abstract away from modelling every detail (Cooper, 2002). However, this has led to claims that models may not generalise to human populations, and that models of language evolution in particular often contain "unrealistic initial conditions" which limit the problem space in non-trivial ways (Bickerton, 2003:86). One of the central goals of this thesis is to examine whether this claim holds up by making explicit attempts to replicate computational results in human populations in order to assess their ecological validity.

3.2.2 Emergence of Natural Human Communication Systems

It is not every day that we get to witness the birth of a new language; the vast majority of us are born into a community with a fully fledged linguistic system firmly in place. The few exceptions to this rule are therefore invaluable, as they give us a unique opportunity to observe the natural emergence of a human communication system. There are two main loci for witnessing such an event: in the formation of pidgin and creole languages (Bickerton, 1981), and in the formation of home-sign (Goldin-Meadows & Mylander, 1998), and full sign languages in the deaf community (Kegl, 1994; Senghas & Coppola, 2001; Senghas *et al.*, 2004; Sandler *et al.*, 2005). One thing that has been emphasised in both studies of protolanguage and emergent sign-languages is the key role that children appear to play in the process.

For instance, Senghas & Coppola (2001) have explicitly focused on the different roles played by adults and children in the development of a relatively young sign language in Nicaragua (NSL). This is a language that has been emerging since the 1970's, when schools were established to educate the country's deaf children, most of whom lived in small isolated communities. Prior to this time, there was no established sign language in Nicaragua, or even a deaf community to speak of. Since this time however several cohorts of deaf people have passed in and out of the school system every year, and a new language has been rapidly emerging. Initially composed of just a few basic signs that were rapidly converged upon, each successive cohort (or generation) of learners has elaborated and systematised the grammar of the emergent language.

The schools contain a mix of children and adults, all of whom have hearing parents. Senghas & Coppola (2001) investigated where the internal structure of NSL was coming from, and found strong evidence to suggest that it was the younger deaf students, and not the adults, who were providing the creative force. They link this back to the fact that children are much better at acquiring language than adults, despite the fact that adults are much better at mastering other complex skills (Newport, 1990). Although Senghas & Coppola (2001) have interpreted this result as showing that there is a qualitative difference in the behaviour of adult and children learners, they do however go on to point out that the status of the evolving language itself also plays a role, stating:

> "Each generation leaves the distinctive mark of their learning process on the model they provide for their children. When children learn a mature language, the mark is a subtle one...Only in cases like this one, when the model is not a mature language, do these language-learning abilities show their transformational, creative capacity." (p 328)

In other words, we only tend to see these creative capacities of children when there is a sparseness of data in the linguistic environment. They do not show up ordinarily during first language acquisition when the linguistic target is already a fully-fledged language. This shows nice parallels with the findings of the iterated learning models discussed earlier, which suggest that languages only adapt when they are culturally transmitted *and* there is some kind of sparsity in the input (Kirby, 2002; Zuidema, 2003; Kirby *et al.*, 2008b).

Although the data deriving from these case-studies tends to be both detailed and directly relevant to language evolution, they do have their limitations. The first has already been mentioned -- they are rare. This makes it difficult to extract robust generalisations. Just as any empirical study needs many data points in order to calculate the size of the effect, we find we need many case-studies in order to be sure we are detecting the common processes underlying the emergence of new languages in general, and not just facts idiosyncratic to the formation of specific languages. The second issue is one of control. Although scientists working on these cases can look at the data and develop hypotheses about what is responsible, they cannot easily go on to test their intuitions by manipulating any of the variables. In most cases, researchers must remain passive observers to the phenomenon at hand, recording what happens but not intervening.

3.2.3 Artificial Language Learning

Another method which has come to the fore in recent years is artificial language learning (ALL). In ALL studies, participants are trained on an artificial language -- usually just a sequence of letter strings generated by a grammar -- exhibiting some set of features controlled by the experimenters (Reber, 1969; Knowlton & Squire, 1994). After training, learners are tested to see what they have acquired, and whether they can recognise novel sequences produced by the same grammar. This has proven to be a powerful method for ascertaining what kinds of structures humans can acquire, and one which is not only useful for studying abilities in adults, but also infants (Saffran *et al.*, 1996; Gomez & Gerken, 2000), and even non-human primates (Fitch & Hauser, 2004)[6].

There have been several ALL studies conducted that have tried to shed light on issues relating to language evolution. Following on from some of the findings obtained from the sign language study described earlier, Hudson-Kam & Newport (2005) used an ALL task to address performance differences between adults and children in terms of how they impose structure by regularising inconsistent inputs. In one study they found that when an artificial language contained irregularity (i.e. a grammatical feature was either consistently present or only present 60% of the time) children were much more likely to impose their own systematic pattern when attempting to reproduce the data than adults were. Although this finding seems to largely support Senghas & Coppola's (2001) claim that children's learning behaviour is categorically different to adults, a follow up study by Hudson-Kam & Newport (2009) complicates the issue somewhat by discovering that there are in fact certain conditions in which adults will regularise and children will not. It seems there are many factors which determine when learners will generalise observed patterns to new data, and when they will not.

The situation becomes more complicated still when we consider a more recent study by Smith & Wonnacott (2010), who show that when adults are engaged in iterated version of Hudson-Kam & Newport's original (2005) study, the languages all evolve to become regular. This study illustrates one of the key findings that comes from my own work in Chapters 4-6. Namely, that the performance of an individual at the beginning of a transmission chain is radically different to the performance of an individual at the end of a transmission chain. Participants in Hudson-Kam & Newport's original study are equivalent to participants in the first generation of Smith & Wonnacott's. What the latter study shows is that although one adult might not regularise, if we have a chain of adults learning from one another, they *will*

[6] These status of these animal studies is currently controversial. More replications of them are required.

regularise. In other words, we cannot predict what the outcome of iterated learning will be on the basis of the performance of the first learner. The significance of this finding will hopefully become clearer after Chapter 4.

Another application of ALL lies in testing specific predictions generated by different language evolution theories (Christiansen, 2000; Ellefson & Christiansen, 2000). This has been used in conjunction with computational simulations (Christiansen & Devlin, 1997; Ellefson & Christiansen, 2000) to investigate whether the brain mechanisms governing the acquisition and processing of language are linguistic or more generally cognitive in nature. For instance, Ellefson & Christiansen (2000) investigate the phenomenon of subjacency. All languages place certain restrictions upon the ordering of words. Violation of any of these restrictions results in sentences which are ungrammatical. The principle of subjacency is an example of one type of restriction which operates on languages. It refers to the fact that when elements undergo movement (for instance, in the formation of wh-questions in English) there are only certain places that a given element is accessible and free to move from (Newmeyer, 1991).

The appearance of seemingly arbitrary subjacency constraints on word movement has been used to motivate the idea that language must be the result of specialised cognitive equipment. In other words, that these restrictions only make sense from a linguistic perspective (Pinker & Bloom, 1990). Ellefson & Christiansen designed an ALL experiment where they presented subjects with grammars that fit either natural or unnatural subjacency patterns, and found that they acquired the natural grammars significantly better. On its own, this might be taken to show that human participants prefer the natural subjacency constraints because those are the ones endorsed by UG. However, they also performed a computational model, using the same data to train a simple recurrent network (see: Elman, 1990). Although the computational agent had no specialised linguistic processing machinery, its performance matched that of human learners. Ellefson & Christiansen (2000) therefore conclude from this that subjacency constraints seen in human languages could have emerged from very general cognitive constraints on sequence learning.

3.2.4 Emergence of Artificial Human Communication Systems

There has recently been renewed interest in studying the emergence and evolution of human communication systems experimentally (e.g. Galantucci, 2005; Garrod *et al*, 2007; Healey *et al*, 2007; Scott-Phillips *et al*, 2009; Selten & Warglien, 2007; Kirby *et al*, 2008a; Theisen *et al.*, 2010). These studies differ from the many experiments investigating human communication that came before (e.g. Garrod & Anderson, 1987; Garrod & Doherty, 1994; Christiansen, 2000; Pickering & Garrod, 2004; Hudson-Kam & Newport, 2005) by the emphasis placed on exploring the emergence of *novel* systems. In other words, these experiments do not start with a system (either natural or designed by the experimenter) in place initially, but let one evolve over the course of the experiment. This provides us with a direct route into understanding how such systems become established (Galantucci, 2005).

It is clear that an experimental approach offers certain advantages over studying these phenomena indirectly via the use of computational and mathematical models, or via naturalistic observation (such as greater experimental manipulation, control, and replicability of results, etc.). Most of these newer experiments looking at the emergence of novel systems share the property of revolving around some kind of communication game. Participants (typically dyads) are given some shared goal or joint task that requires them to co-ordinate their actions in some way. The only way in which to do this is to interactively construct a communication system together, using whatever medium is provided.

For instance, in Selten & Warglien (2007) pairs of participants are given a repertoire of available symbols, each with different sending costs, and instructed to converge upon a set of economical signals to identify different pictures. In Galantucci (2005) pairs of participants must coordinate their actions in a 2D game-world by communicating with one another using a novel graphical medium, which prevents the use of common symbols or pictorial representations, forcing them to develop a new system of their own. In Healey *et al* (2007) pairs of participants (and later on, interacting groups) collaborate together using a virtual whiteboard, drawing images to identify different pieces of music. Similarly, Garrod *et al* (2007) encourage participants to depict various concepts (such as commonly known people, places, objects, and more abstract concepts such as 'poverty') using images in such a way that a fellow participant could identify them. In a slightly different twist, Scott-Phillips *et al.* (2009) have an experimental set-up in which they do not even provide a dedicated channel for communication to take place in: given a task which requires two players to coordinate their actions, the only solution is to create one by using the movements of the players' avatars in the game environment as signals.

The fact that convergence does not come easily to participants in these experiments (most fail to agree on a system, and fewer still go on to develop one with structure) highlights the fact that the underlying processes responsible are not trivial. This is perhaps surprising given that we assume participants could easily invent a workable system on their own. In fact, Scott-Phillips *et al* (2009) find that reported reasons for failure often centre around an inability to convey a system to their partner rather than an inability to individually construct one in the first place. Conversely, Selten & Warglien (2007) showed that the chances of developing a successful system are massively increased when one player finds a way to take control and impose their invented system upon the other. This raises the interesting question of what kind of design process we think is responsible for the emergence of structure in natural language -- is it one which is wholly reliant on the ingenuity and design skills of its users, or is there some other force at work?

Although these studies all show that humans are adept at constructing novel communication systems, the next section argues that many linguistic changes are not ‘designed’ by individuals in that manner. Rather, much of the structure present in human language is indicative of apparent design *without* a designer.

3.3 Design without a designer

For centuries philosophers and linguists have debated the origins of linguistic structure and how languages change. One of the central mysteries involves identifying the source of those changes and innovations that lead to increasing structure. The intuitive answer is of course us, the speakers of language. Yet whilst languages change and evolve as a result of differential patterns of usage among speakers, they do not do so as a result of any intentional design on the part of an individual. As Keller (1994) points out, we cannot analyse a historical change like the shift in word ordering from Object-Verb to Verb-Object in Middle English, and come to the conclusion that it is an instance of human design.

Keller refers to events like this as phenomena of the third kind - grouping together things that are neither man-made (artefactual) nor entirely natural, but which are instead “the result of human actions but not the goal of their intentions” (Keller, 1994:56). He argues that as most language changes are of this type, we need to invoke an ‘invisible hand’ explanation for language, adopting the metaphor proposed by the economist and philosopher Adam Smith to explain how locally self-serving actions of individual investors can unexpectedly lead to group-level prosperity. If this hypothesis is correct, it is only through developing an understanding of how apparent design emerges *without* a designer that we can hope to discover the origins of linguistic structure.

Croft (2000) makes a similar three-way distinction between types of causal mechanisms involved in language change to that proposed by Keller (1994). On one hand, we have TELEOLOGICAL explanations, which are invoked “when a speaker is claimed to innovate in order to alter the linguistic system in some way...the linguistic system is designed (by the speaker) to have the structure it does, and to change, as it does" Croft (2000:64). This corresponds with what Keller calls man-made. Like Keller, Croft concludes that this is not a mechanism that operates in language change. Next we have INTENTIONAL explanations, where "the speaker is aiming towards some other goal in language use, and produces an innovation in the process" Croft (2000:64). This corresponds to Keller's phenomenon of the third kind. We have seen evidence of this kind of mechanism at work in the experiments described in §3.2.4. The final kind of causal mechanism in language change involve NONINTENTIONAL explanations, where "[t]he language change is not even an intended means to achieve some other goal of the speaker. It is simply a change that just happens as a consequence of the act of production (and in some theories, also comprehension) of an utterance" (Croft, 2000:65)[7]. It is this kind of mechanism that I would like to investigate with the experiments in Chapters 4-6.

For Keller (1994), who views language change as a special instance of sociocultural change, explaining the properties of language inevitably requires seeing it as a product of cultural evolution. Although Keller primarily restricts his investigations to language alone, the invisible hand phenomenon is also at work in many other domains, for instance, in how crowds of people self-organise into the optimal spatial configuration for viewing performers. However, it is certainly not the case that *every* instance of cultural evolution requires an intentional or a nonintentional explanation. If we look outwith human communication, we find that many examples of culturally transmitted behaviours, such as tool-making and the kinds of incremental innovations we find in technological developments (Basalla, 1988; Petroski, 1992; Ziman, 2000), do seem to be directed and guided by human

[7] This corresponds to what Keller (1994) calls 'natural kinds'. Keller does not place as much emphasis on these kinds of changes within language, which is one of the reasons why I will be using Croft's definitions for the rest of the thesis.

intentions - they do require teleological explanations[8]. In that sense, we can see Croft's (2000) three causal mechanisms as operating more generally within cultural evolution.

For some commentators (e.g. Hallpike, 1986; Pinker, 1997; Benton, 2000; Bryant, 2004), this teleological or goal-directed aspect is precisely what causes analogies between biological and cultural evolution to breakdown completely (Mesoudi, 2008). Instead of perceiving this as an either-or debate (in which cultural evolution either proceeds via intelligent human design or some blind evolutionary process), Dennett & McKay (2006) encourage us to think of cultural change as: "a continuum from intelligent, *mindful* evolution through to oblivious, *mindless* evolution" (italics original). They go on to claim that:

> "in cultural evolution...there are undeniable cases of cultural features that evolve by Darwinian processes without any need to invoke authors, designers, or other intelligent creators. Most obviously, languages - words and pronunciation and grammatical features - evolve without any *need* for grammarians, deliberate coiners, or other foresighted guardians of these cultural items." (p. 353).

So this brings us back to our central question - if some aspects of linguistic structure are led by this invisible hand, or are in fact completely nonintentional as Croft defines it, is it possible to capture this phenomenon and investigate it in the laboratory? It could be argued that, in a sense, we have already seen the invisible hand at work in some of the studies discussed in §3.2.4. Whilst the interactions between participants do involve some degree of reasoning and purposeful design, participants' intentions were to cooperate to solve the task. Although they were all

---

[8] Sometimes it is uncertain whether or not the inventors themselves can anticipate the eventual usage of the object to which they contribute some design feature. In this case, we are back to describing these changes as intentional. The original innovator may have intended to make a modification to improve the way in which a stone tool cracks nuts, but another observer may see that object, and believe that its proper function is as a spear-head. The creation of a new spear-head was not what the modification was designed for, but the modification itself was nevertheless intended.

consciously aware that they needed to find a way to communicate with their partner, the negotiation process which allowed the basic communication systems used by different participants to become aligned with one another and become an established convention is also a complex dynamic system at work. As such, it has invisible hands of its very own; shaping, guiding and prompting structure into being. This notion would help to explain why the creation of a successful system is never guaranteed in these studies, in spite of the fact that an individual acting alone given explicit instructions to design a way to communicate, could easily invent a system fit for purpose.

However, if we genuinely want to explore the nonintentional end of the scale, we need to design an experiment where participants are not given the explicit task to communicate. Isolating exactly which elements arise through intentional design, and which through these more subtle and hidden forces, may well prove to be impossible in *any* experiment involving human participants. However, that should not prevent us from trying.

3.4 The Current Framework: The Human Iterated Learning Model

This section lays out the experimental framework that will be used for the rest of the thesis. It begins by clarifying a recent point of confusion amongst researchers about what iterated learning really consists of, before describing some early work that was done to investigate language change that bears many similarities with the suggested framework. Finally it looks in some detail at ways in which we can analyse the results of our experiments.

3.4.1 Putting Iterated Learning in Context

There has been something implicit in all of the discussion so far that should be made explicit at this point. *All the previous studies of language emergence in the laboratory are instances of iterated learning*. What makes this experimental framework different to these other approaches is not that it involves iterated learning, and the others involve some other transmission mechanism. Instead, the difference lies in two factors: (1) population structure, and (2) the focus on nonintentional emergence of structure. This has caused some confusion in the literature of late, mostly as a result of the fact that the majority of researchers who have used the term 'iterated learning' before, in both linguistic and non-linguistic domains, have demonstrated it by using simulations (Kirby, 2000; Kirby, 2001; Kirby & Hurford, 2002; Zuidema, 2003) or experiments (Kirby *et al.*, 2008a; Griffiths *et al.*, 2008; Smith & Wonnacott, 2010) that involve linear transmission chains of learners. This has in turn led other researchers to attempt to make contrasts between approaches which, strictly speaking, should not be made.

For instance, in a recent paper Garrod *et al.* (2010:33) state that: "One influential model assumes an evolutionary principle analogous to iterated learning in which the language is transmitted vertically down generations of speakers". Using this definition of iterated learning, Garrod *et al.* (2010) go on to contrast two different experimental conditions: iterated learning and social coordination. The contrast that was actually being made here was between linear transmission vs. closed group population designs. Of course there is nothing wrong with researchers redefining terms as they see fit. However, this particular example is dangerous as it implies there is a difference between the two conditions in terms of the mechanisms of transmission at work, rather than a difference in population structure.

3.4.2 General Methodological Framework

This section will outline the general methodology for conducting iterated learning experiments to investigate the nonintentional emergence of language. The idea is that learning something about the way in which artificial languages are culturally transmitted in the laboratory can tell us something about the way in which natural languages are culturally transmitted in real populations.

The general method involves each participant learning a small artificial ('alien') language composed of a finite set of meanings (pictures) that are paired with signals (strings of letters, or possibly sounds). These languages need not be particularly large. In the experiments described later on in chapters 4 and 5 there were just 27 meaning-signal pairs in total. Once a participant has acquired the artificial language, they are tested and their answers used to provide the training input to the next participant, who forms another 'generation' in the chain. This process repeats until the desired number of generations is reached. Throughout, participants are asked only to reproduce the language as accurately as they can; the source of their training data is not revealed, and they have no way of knowing the experiment is investigating the emergence of language.

Training, Testing and Transmission

There are three distinct phases involved: training, testing, and transmission. During the **training phase**, participants are shown a picture from the set, alongside the signal string it is paired with, and informed that this is the way in which the alien would describe that image in its own language. The task is to learn the descriptions associated with each image to the best of their abilities. Training occurs via a computer program, which randomises the order in which each signal-meaning pair is presented, ensures that all training items are seen, and controls the length of time each training item is shown. The key variables to consider here are the amount of training each participant receives (i.e. the number of rounds of training they are given), whether this training occurs in one continuous session or in blocks, and whether training blocks are structured in some way or randomised.

Following a series of pilot studies conducted during my MSc (Cornish, 2006) it was decided that training would be conducted over three blocks, with an obligatory practice test and an optional two-minute break in between. Each training item appeared twice during each block, so six times in total over the course of a learner's training session. Training items were presented in two parts: first the signal would appear alone for 1000ms, then the meaning would be shown alongside it for a further 5000ms.

Once training is complete, we move onto the **testing phase**, where participants are shown each picture in turn and instructed to supply the missing description. The final test can be (and in the experiments presented later on in Chapters 4 and 5, in fact were) preceded by a series of practice tests in between training blocks, which introduces the possibility of some indirect feedback being provided to facilitate learning: participants were given a limited opportunity to correct themselves over the intervening practice tests, as well as giving them the chance to become familiar with the testing procedure ahead of the final test. In the experiments presented here, the practice tests involved participants being presented with just a subset of 14 of the meanings and being asked to provide the correct description[9]. Following the third and final block of training, the remaining 13 items were appended to this set of 14, ensuring that descriptions were collected for all 27 meanings in the final test.

These responses from the last round of testing are then used to generate a new set of training stimuli for the next generation during the **transmission phase**. It is during this final stage, which happens 'offline' after the participant has left, that some of the most interesting parameters can be explored, including the transmission bottleneck.

[9] Depending on the exact condition of the experiment, half of these items may have been ones which they had seen earlier in training, and half may have been novel.

One of the advantages of the iterated language learning methodology is that it allows us to test very specific hypotheses about what occurs during language transmission by giving us complete control over what gets passed on. It is this aspect that affords iterated language learning more simulation-like qualities than is typical in non-iterated artificial language experiments.

For instance, if we wished to test the hypothesis that a preference for shorter strings led to compositional structure, during the transmission phase we could artificially select only those strings that met some (possibly dynamic) string-length threshold and ensure that only these items were propagated to the next generation[10]. By examining the resulting languages that arise from this process of artificial selection we can determine whether this hypothesis is valid. In this case we are running the procedure like a simulation. We build in a condition to see what the future outcome is, and can then refine our intuitions as a result. Alternatively, if we wish to test the hypothesis that human learners actually *have* a bias towards producing shorter strings, we can just run the experiment without any such manipulations and examine the average length of strings at the end of the chain. In this case, we are using the methodology to experimentally test whether such a bias currently exists or not. Both strategies can be useful depending on the questions one wants to answer.

To summarise, the procedure implemented in most of the experiments described in Chapters 4 and 5 was as follows: (1) participants are given verbal and written instructions asking them to learn the alien descriptions for a series of images; (2) three blocks of training occur; (3) final responses are gathered in the last test; (4) participants were debriefed; (5) the recorded output from each participant was processed ready for transmission to the next learner. During every block of training,

[10] It should be remembered that studying processes of artificial selection (e.g Mendel's peas, the selective breeding programs employed by farmers, etc.) were what led to the breakthroughs in understanding how biological evolution worked. One of the points being argued here is that similar tactics of studying artificial selection in language and other culturally transmitted behaviours can lead to similar advances in understanding cultural evolution. This is consistent with the agenda laid out in Mesoudi et al. (2006).

each training item is seen twice. Participants are then tested on roughly half (14) of the items, and given an optional 2 minute break before the next block of training commences. This sequence is depicted in Figure 3.1.

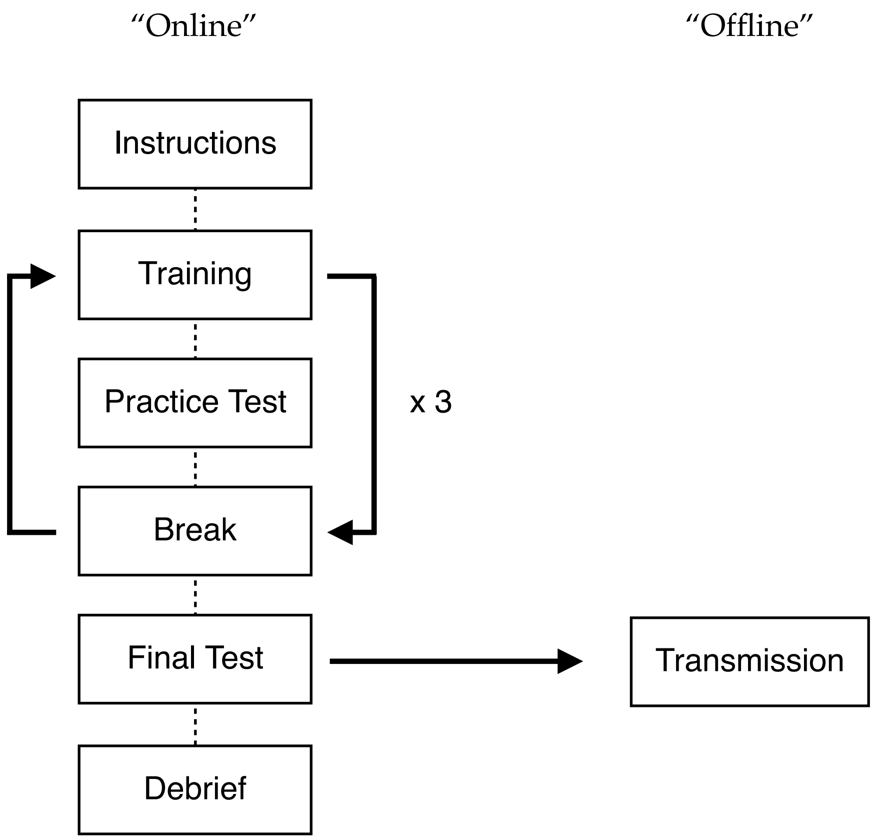


Figure 3.1: The training-testing-transmission procedure for experiments I-IV. Training and testing occurs 'online' (when participants are in the laboratory) whilst transmission occurs 'offline' (after participants have left). Each experiment involves three blocks of training and testing. Only the output from the final training cycle is processed ready for transmission to a new learner.

Generating Initial Languages

The experimental procedure is only one of the considerations that need to be kept in mind. One obvious factor we have yet to mention is how we begin this process. It is clear that the first participant needs a language to learn. There are several manipulations we can make here, which are again dependent on the kinds of questions we are interested in. For instance, if we wish to know whether a particular structural system can be stably transmitted, then we should give that system to the first participant and monitor whether it changes as a result of iterated learning. If however, we are interested in learning something about how linguistic structure emerges, we cannot initialise the chains with a fully structured system. Instead, we can use randomly generated signals. A simple method for constructing these is by concatenating CV syllables (drawn from a large but finite set) to form longer strings. This produces a set of signal strings which, whilst containing some regularities owing to the fact that they are constructed from a finite syllable set, is still highly unstructured with regard to the meanings.

The Meaning Space

Further consideration must be paid to the design of the meaning-space - or rather, the stimuli we use to depict the meaning-space. Meaning-spaces themselves can be structured or unstructured, reflecting regularities and co-occurrences in the real world, or a controlled and simplified world of our choosing. In all of the studies discussed later, the pictures come from a small and highly structured meaning space consisting of three different dimensions (motion[11], colour and shape), each of which contains three different variables (e.g. bouncing, straight and spiralling; black, blue and red; circle, square and triangle). This 3x3x3 design yields a total of 27 different possible combinations. Some examples of these meanings are shown in Figure 3.2 below.

[11] In the actual experiments presented in Chapters 4 and 5, motion was indicated by dotted lines. This can be seen in Fig. 3.2.

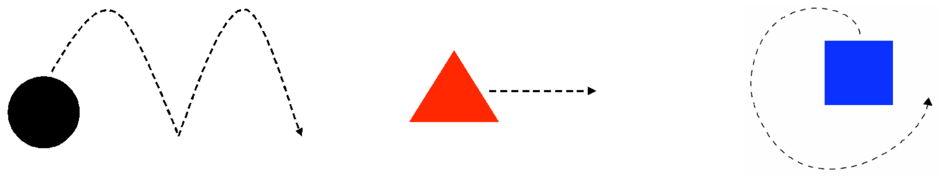


Figure 3.2: Examples of the images used to depict structured meanings in Experiments I-IV. Each meaning varies in terms of motion, shape and colour. These examples show a bouncing black circle, a horizontally moving triangle, and a spiralling blue square.

Population Structure

The population structure can be manipulated in a variety of different ways. Not only is it possible to control the size of the population, but also the network structure (i.e. who talks to who). Since there are so many possible configurations, it makes sense for us to look at the simplest possible population structure first: a linear transmission chain, with just one learner at every generation. It is important to remember however that we can also implement closed-group and replacement designs, or in fact, have one learner receiving their own input back in a disguised manner, as in Griffiths, Christian & Kalish's (2008) exploration of category learning[12].

3.4.3 Measuring Structure and Learnability

The next chapter reports the results of the first two experiments. Before we begin, it is perhaps worth spending a moment considering how we are to analyse the data. Given our hypotheses, we need to know two things. Firstly, whether the languages are evolving to contain more structure, and secondly, whether they are becoming easier to learn. In simulations, modellers have free access to the grammars being constructed by agents over the course of each run, which makes it relatively

[12] See Winters (2009) or Line (2010) for some examples of population manipulations that have been tried within human iterated language learning.

straightforward to describe and make comparisons of the systems at different stages in their evolution. If we want to know whether the language at the final generation is structured, we can simply inspect the internal representation of it in the 'mind' of the agent and find out[13]. Although our use of human participants rules out such a direct approach, we still have plenty of resources available to us, most notably the forms of the languages themselves, to enable us to objectively judge the matter.

The issue of learnability is relatively straightforward. In short, *a language is learnable to the extent that it is transmitted faithfully without error*. In order to measure error in transmission, we need only to find a way of calculating the amount of change between different languages. So what determines whether a language is structured or not? We should remember here that we are dealing with a simplified definition of what a language is. In these experiments, a language is simply a mapping between meanings and signals. With that in mind, a language can be said to be structured if that mapping between the different levels (meanings and signals) is itself structure-preserving. In other words, *a language is structured if similar signals get reliably mapped onto similar meanings*[14]. We therefore need a measure that can tell us whether there is a correlation between items that are similar in one dimension (meanings), and items that are similar in another (signals). The rest of this section explores techniques that allow us to do that, starting by examining methods for quantifying the distance between signals and meanings.

---

[13] This is at least true for symbolic models (e.g. the ILM described in section 3.1 of Kirby & Hurford (2001)). Although it is also possible to access the internal states of agents in connectionist models it is not always particularly meaningful to. As Russell & Norvig (1995: 584) explain, connectionist models "are essentially black boxes". Whilst some modelers have used mathematical techniques such as hierarchical clustering (Elman, 1990) and Principal Components Analysis (Elman, 1991) to try to understand what is going on, these methods only allow abstract comparisons between network states. Although connectionist models produce rule-like behaviour, they do not represent rules locally (Bechtel & Abrahamsen, 2002).

[14] Obviously the origin of the structure in a language is going to come largely from the structure in the world. This is necessarily the case in these experiments given that we are providing our participants with a highly structured meaning-space. The final experiment in chapter 5 will address this issue in a different way, but for now we are simply interested in whether signals can come to reflect useful structure present in meanings.

Distance Metrics

How do we begin to go about measuring similarity in our language domains? One of the advantages of running experiments involving a fixed set of meanings is that it should make the task of constructing a simple measure of language structure much easier (Galantucci & Garrod, 2010). The fact that the meanings are predefined, and can be easily decomposed into features with different values means that they lend themselves nicely to being defined spatially. As suggested by Brighton *et al.* (2005), we can view each meaning as a vector in some Euclidean space. Each vector is defined by two components: the feature of the meaning (in this case, colour, shape or motion) and the value of that feature (i.e. 'blue'). These dimensions reflect and define the meaning space: so a 3x3x2 meaning space consists of 18 meanings varying along three features, the first two of which have three values, and the last having only two values; whereas a 5x5 meaning space consists of 25 meanings, that vary along five features and five values.

Because the meanings in our experiment vary consistently in terms of the number of features and values, we can use Hamming Distance (HD) for the meaning-space (Brighton *et al.*, 2005). This is a standard metric from information theory that looks at the number of substitutions required to convert one string into another (metrics like this are commonly referred to as *edit distances*, as they involve computing the number of changes required to get from state *t* to state *t+1*). In this case of our experiment, two meanings are compared against one another and for each feature value (motion, shape and colour) that differs between the two, a point is awarded. So for instance, a bouncing black square and a bouncing black triangle differ in a single feature, and therefore have an HD of 1, whereas a bouncing black square and a horizontal red circle differ in all features, so have an HD of 3.

Given the precedent already set for using Hamming distance to measure similarity between meanings in this context (Brighton *et al.*, 2005), there are no problems with adopting it for use in the experiments. However, there are alternative ways to measure similarity between meanings, that we will briefly consider here. One of the reasons for adopting HD as a metric is that we know the exact features and values of our meanings in advance, and have a relatively simple semantic structure. When the exact semantic structure is unknown or high-dimensional, other techniques must be used.

For instance, Shillcock *et al.* (2001) examined the level of systematicity between the forms and meanings of 1733 monosyllabic and monomorphemic English words, taken from the British National Corpus[15]. In order to measure the semantic distances between the different word meanings they first had to examine the lexical co-occurrences of these words in the entire 100 million-word corpus. Using the vector-space method presented by Lund & Burgess (1996), Shillcock *et al.* (2001) used this co-occurrence data to construct a semantic space containing some 500 dimensions. Each point in this high-dimensional vector-space represented a meaning, and the distance between any two points could be calculated using the angles between these vector points.

Lexical cooccurrence matrices are commonly used in the construction of semantic spaces in computational linguistics because being automatically induced, they avoid the problem of relying on the experimenter identifying the correct dimensions - a task which becomes exponentially more difficult to calculate as the number of meanings increases (Jurafsky & Martin, 2000). They also capture the intuitive idea that the meaning of a word is (at least somewhat) determined by the linguistic contexts in which it occurs (Tamariz, 2008). Lund & Burgess' (1996) vector-space method is closely related to an approach known as Latent Semantic Analysis (LSA), developed by Landauer & Dumais (1997) as a more general solution to what has

[15] See also Tamariz (2008) who used the same basic technique to investigate systematicity between forms and meanings in a corpus of spoken Spanish words.

come to be known as Plato's problem: namely, how do we come to know so much, given so little experience?[16] LSA works on the idea that:

> " some domains of knowledge contain vast numbers of weak interrelations that, if properly exploited, can greatly amplify learning by a process of inference....[T]he choice of the correct dimensionality in which to represent similarity between objects and events, can sometimes, in particular in learning about the similarity of the meanings of words, produce sufficient enhancement of knowledge to bridge the gap between the information available in local contiguity and what people know after large amounts of experience." (Landauer & Dumais, 1997:211)

By using the statistical properties of contextual co-occurrence, and very general induction mechanisms, Landauer & Dumais (1997) built a model which could acquire knowledge of English vocabulary from noisy internet chat forums at a similar rate to school children. This happened despite the fact that the model had no prior linguistic or perceptual similarity knowledge. The idea of measuring 'similarity' may actually be more than a useful metric for our research purposes. Although we are only interested in calculating similarity between signals and meanings to determine whether the languages in our experiments are being faithfully acquired and more structured, it turns out that this could be something real learners are also tracking during acquisition.

Returning to the topic at hand however, just as we can consider using edit distance to compare meanings, so too can we use edit distance to compare signals. One potential complication with comparing edit distances for signal strings lies in the fact that the string lengths can vary. Instead of Hamming distance then, which relies on symmetrical lengths, a better metric is Levenshtein Distance (LD). This calculates not just the number of substitutions to turn one string into another, but also handles

[16] The term "Plato's problem" has been used by Chomsky (e.g. 1991) to refer explicitly to poverty of stimulus arguments in language acquisition. However, the term should be understood as being applicable to a wider set of induction problems than just those relating to language acquisition.

insertions and deletions as well (Levenshtein, 1966). For example if we wanted to compare the similarity between two strings, **kopafilo** and **kapilo**, we would calculate the most efficient way of turning one into the other: in this instance there is one substitution (o to a) and two deletions (a and f), resulting in a Levenshtein Distance of 3. This figure can be normalised to give a value between 0 and 1 by simply dividing the LD by the length of the longest string (Brighton *et al.*, 2005) - in this case giving us a value of 0.375.

Once again we find that this is just one of many different ways in which we could measure distance between signals. Strictly speaking, if we were using spoken signals we should weigh the edit distances according to how frequently we observe that kind of edit (read: error) in a given phonetic environment. For instance, given the fact that unvoiced sounds have a greater tendency to become voiced if they appear intervocalically, we should perhaps give less weight to this kind of change as opposed to a more unusual one. In addition to the standard version of LD described here, Kessler (2005) reviews different techniques for measuring phonological similarity and describes versions of LD with different weights given to reflect the greater salience of certain types of edits. There are two reasons for not using any of these more sophisticated versions of LD in the present study however.

The first is that because the signals in the experiment were visual, and not phonetically transparent[17], we have no idea if changes are likely to be the result of (a) typological mistakes, (b) phonological mis-parsings, or (c) combinations of both of these. Secondly, estimating these weightings relies upon native speaker judgments. There are no ‘universal’ patterns - everything is determined by the particular phonological structure of the language in question. As the signals used in these iterated language learning studies are artificially constructed, we have no native speakers. Even if we restrict our studies to only include monolingual English

[17] Given the signal string **maciro** we cannot be sure whether participants will phonologically parse that as [məkiɹɔ], [mɑkiro], [maʃiɹo] or something else.

speakers, and treat the alien signals as pseudo-words, there are further impracticalities that arise.

As an illustration, Tamariz (2008) measured systematicity between forms and meanings in the Spanish lexicon using a large corpus of transcribed spoken utterances. In order to generate a measure of phonological similarity, a previous empirical study was run (Tamariz, 2005) to collect similarity judgements from native speakers. Using these perceptual judgements, a set of parameters can be devised and applied when comparing two strings. However, in order to do this, not only did string length have to be controlled for, but the syllable structure as well. This meant restricting investigation of similarity only to words conforming to the following structures: CVCV, CVCCV, or CVCVCV. Given the fact that we cannot restrict the output that each participant produces after training in any way, generating perceptual weightings for string similarity would be an exhaustive task in its own right.

Detecting structure within a language - The Mantel Test

Once a suitable set of metrics has been found for determining signal distances and meaning distances, how do we use those to judge the amount of structure between them? Again, we find that a number of different approaches have been taken in the literature (Shillcock *et al.,* 2001; Brighton *et al.*, 2005; Ellison & Kirby, 2006; Tamariz, 2005). Whilst on the surface all of these measures appear quite different, they are in fact all just variants of a test proposed by Mantel (1967), which has been used more commonly to explore patterns of correlations between different distance metrics within ecology (Sokal & Rohlf, 1995). The rest of this section will explain how this test works by using a toy example from that domain.

Essentially Mantel's test assesses the correlation between two symmetrical matrices, each cell of which contains the distance between an object and every other object in the set. Imagine we were interested in whether species with similar genes had similar geographical distributions. The first matrix would therefore contain all of the genetic distances between all possible species in the study, and the second would contain all geographical distances between those same species. This is illustrated using hypothetical data in example 3.1, where {**a,b,c**} are three different species, and numbers represent some notional distance in the relevant domain.

(3.1)

geographical distance

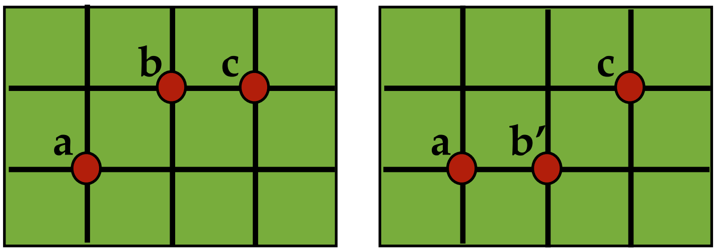
	a	**b**	**c**
a	0	2	2.24
b	2	0	1
c	2.24	1	0

genetic distance

	a	**b**	**c**
a	0	4	3
b	4	0	2
c	3	2	0

Typically when we are wanting to see whether two variables like this co-vary in an interesting way we can simply perform a statistical test to determine the strength and direction of any correlation, and the degree of confidence we have about that correlation being genuine. We cannot do this here however. The problem with making a straightforward correlation between the two matrices is that distances, by their very nature, are not independent from one another. In a matrix containing n objects, if you could imagine moving one of them slightly, $n-1$ distances would also change as a direct result.

To illustrate this more clearly, consider Figure 3.3 below. The space on Fig. 3.3.left depicts the space represented in the original geographical distance matrix outlined in example 3.1. If we imagine moving datapoint **b** slightly (Fig 3.3.centre), from position [2,2] to [2,1], we have not only changed the matrix cells for that one object, but also all the distances from that object to all of the others (Fig 3.3.right). It has moved closer to **a** and further away from **c**.



geographical distance

	a	b′	c
a	0	1	2.24
b′	1	0	1.41
c	2.24	1.41	0

*Figure 3.3: A pictorial representation of the geographical space depicted in example 3.1 (left) and the consequence of moving one of the elements (centre). This figure demonstrates the fact that distances are not independent of each other. Not only has the location of **b** changed (to **b′**) but all of the distances between **b′** and every other location, as shown in the new geographical distance matrix (right).*

This lack of independence means we cannot rely on standard parametric tests to show significance. Mantel's (1967) solution was to perform a Monte-Carlo (or permutation) test on the two matrices in order to calculate significance that way. The way this works is as follows. First we go ahead and calculate the correlation anyway. The exact test we use depends on the nature of the data and the distance metrics we have decided to use. Given that we have adopted the same distance metrics used in Brighton *et al.* (2005) it makes sense to use the correlation measure that they use: Pearson's product-moment coefficient.

Once we have this coefficient, this becomes our veridical score. Next we take one of the two datasets, and we shuffle the order of elements within it. This destroys the veridical mapping between our two distance measures but preserves the actual data, effectively asking the question of what would happen if the exact mapping between the two was unimportant. In terms of the matrix itself, it has the effect of randomly shuffling its rows and columns. Example 3.2 shows what this shuffling procedure looks like when applied to the genetic distance matrix. We then recalculate the correlation on this new randomly aligned data, and judge whether it it the same or greater than that observed in the veridical.

(3.2)

geographical distance	**a**	**b**	**c**
a	0	2	2.24
b	2	0	1
c	2.24	1	0

genetic distance	**a**	**c**	**b**
b	4	2	0
a	0	3	4
c	2	0	1

Crucially we must reshuffle this matrix thousands of times to construct the frequency distribution of scores[18]. This gives us two things. Firstly we can now calculate a level of statistical significance (the *p*-value) in an intuitive and safe (in terms of our data violating the independence assumption) manner by simply counting the percentage of times a correlation is discovered that is equal-or-greater-than the veridical. Secondly using the information derived from the frequency distribution (basically the mean and standard deviation) we can standardise the veridical score to derive a *z*-score. Importantly, the *z*-score, unlike the actual veridical correlation score, can be used to compare observations across different frequency distributions. This may not be important for our toy example but it will certainly be important for comparing our languages, as the data at each generation will have different distributional properties depending on the exact forms and mappings that it contains.

Judging learnability across languages - Transmission Error

The Mantel test examines the structural properties of languages *within* generations, but we also need a measure to assess the similarity of languages *across* generations. For this, we can use a distance metric we have already encountered - the Levenshtein Distance. Whereas previously we used it to compare each signal to every other signal within a language, this time we will use it to derive one number

[18] Of course, with this toy example where n=3, the matrix is so small that we would easily discover every possible permutation quite quickly. However, as the number of possible permutations increases factorially with n, a large number of randomisations is preferred in order to sample as many of these as possible. In the studies presented in the next three chapters, the results were drawn from 10,000 randomised samplings.

that tells us how similar a language is to another language. Taking each meaning, *m*, in turn, we calculate the normalised Levenshtein Distance (nLD) for the signals s_m and s_m' in the two languages we are comparing (Brighton *et al.*, 2005). We then simply average this score over all 27 meanings, giving us a number that varies between 0 and 1 which tells us the average amount of transmission error that there is for each signal in the language. A figure of 0 means that there was no error at all during transmission. In other words, the language was perfectly learnable. A figure of 1 would imply that none of the signals were reproduced faithfully at all. In practice, hitting either extreme of this measure is difficult, as having a single misplaced letter detracts from the maximum score, and even chance correspondences produce scores greater than zero.

The measures for calculating transmission error and structure that have been described here and chosen to be the standard measures used for the rest of this thesis (i.e., nLD, and the Mantel test using HD and LD) have been selected for two reasons. Firstly, on the basis of their generality -- they both appear well suited to detecting all kinds of structure and similarity. Secondly, they have been selected because they are well-understood in the context of iterated language learning with simple meaning-spaces (Brighton *et al.*, 2005). Note however that there are additional measures of structure and learnability available which tend to be better suited for measuring certain types of structure than others. We will see some of these more specialised measures in Chapter 5, and again in Chapter 6.

3.5 Summary

This chapter began by exploring some of the reasons why language evolution has proven to be such a difficult subject to study. Although we have abundant evolutionary end-points of the process around us in the form of modern day languages, there are very few uncontroversial clues to be gained from the fossil record or archaeological data to support our theorising. Next we considered other forms of evidence that could shed light specifically on the processes of cultural evolution. We then looked at some of the pros and cons of current empirical approaches.

Starting with computational and mathematical studies of the origins of language, we noted that many of these models had been criticised in terms of their ecological validity, or for the kinds of simplifying assumptions they make. Next we explored the natural emergence of human communication systems, such as new sign languages. Whilst these case studies do provide us with a wealth of highly relevant data, they do suffer the downside of being very rare. In addition, although researchers can record and monitor the development of these new systems, they cannot intervene or manipulate the process of emergence in order to test specific hypotheses. The next empirical strand we focused on was artificial language learning. This technique has been used both to test specific hypotheses related to language evolution (for instance, to ascertain whether human performance matches the performance of computational models), as well as to study processes identified by researchers working with sign languages and creoles as being relevant.

The final set of empirical studies we examined concerned those which investigated the emergence of novel communication systems in the laboratory. This, as with the artificial language learning studies, has the advantage of providing us with complete experimental control. One thing that was noted about the majority of these studies is that the participants involved are always consciously aware that they need to communicate with a partner or group members, and in some cases, take deliberate steps to try to invent a system to allow them to do. This led us to consider what kind of process we think is responsible for the appearance of design in language.

Here two complementary theoretical frameworks were presented, with which to think about language change. Both frameworks agreed that language was not the result of purposeful design on the part of individuals: it might emerge as a result of human actions, but those actions were not deliberately intended. There were two ways in which language could end up having ‘design without a designer’ - either a language user could intend to achieve some kind of goal with their language use and inadvertently produce an innovation at the same time (this type of causal mechanism was defined as INTENTIONAL); or else a language user could have no higher goal in mind, but make a change as a consequence of the act of production or comprehension (this was referred to as NONINTENTIONAL). Whilst many studies have investigated intentional processes of language emergence, none were found to focus exclusively on nonintentional processes. This observation motivated the design of the experimental framework which was described in the rest of the chapter.

Chapter Four

Language Adapts to be Learnable

4.1. Bottlenecks on Transmission

As discussed in the previous chapter, one of the most crucial parameters within the ILM is the size of the transmission bottleneck. Changing the size of this parameter changes the dynamics of iterated learning considerably. If learners are exposed to the entire set of meaning-signal pairs, the initially holistic system is able to be entirely learned by rote and never changes. If learners are only exposed to a tiny fraction of meaning-signal pairs, the system never becomes stable. It is only when the bottleneck is neither too large, nor too small, that we begin to see systems emerge that are compositionally structured and stable (Kirby, 2000). Given its relative importance in terms of explaining the emergence of language-like structure then, it is surprising that the bottleneck itself has not received that much theoretical scrutiny in the modelling literature.

This chapter begins by exploring the notion of the bottleneck in more detail. It starts with the observation that simulations of iterated learning have tended to model the transmission bottleneck as somehow external to the agent, a distributional fact of the environment rather than anything to do with the way the agent processes the training data internally. It then goes on to present the results of two iterated language learning experiments that investigate what, if any, difference these bottlenecks make to the way in which languages emerge[1]. The first experiment follows the approach used in the simulations, investigating what happens when cultural evolution is driven by a pressure to generalise to novel stimuli present in the environment, whereas the second examines what occurs when pressure comes from a more naturalistic memory constraint internal to the learner. It will be shown that in both cases the languages are adapting under pressures for greater learnability, and consequently, become more structured over time. However, neither produces systems which are optimal for communication.

4.1.1 A Closer Look at the Bottleneck

The idea of a bottleneck in the transmission process is not controversial. One of the principal challenges facing any account of first language acquisition, or indeed, any general theory of linguistics, is to explain how it is that the child converges on the correct grammar for his or her language based on the highly variable and finite exposure to that language that they receive (see Fig 2.2 for characteristics of the linguistic input available to the child). We all arrive at the acquisition process having encountered only a small subset of the possible words and utterances in our language, and yet somehow we manage to negotiate the tricky path towards comprehending it in its entirety.

Whilst the jury is still very much out on the issue of whether the quality of the linguistic input available to the learner is really so impoverished as to necessitate a helping hand in the form of innate linguistic knowledge or not (see §2.2.1 and §2.2.2 for this debate), it is clear at least that language is still somehow being acquired

[1] The experimental results reported in this chapter have appeared in several publications: most notably in Cornish (2010), but also in Cornish, Tamariz & Kirby (2009); Kirby, Cornish & Smith (2008); and Kirby, Smith & Cornish (2008).

despite differences between individuals' data-exposure histories (to use the terminology of Pullum & Scholz, 2002). So what causes these differences in data-exposure, and just how much do they influence language? To answer this question, we need to look in more detail at how the bottleneck is working to constrain the process.

4.1.2 Different Types of Bottleneck

Within the simulations, transmission constraints have most typically been operationalised as the amount of training data given to each learner agent - what Hurford (2002) calls a **semantic bottleneck**. Given that the number of possible conveyable meanings in the models is usually large but finite[2], this bottleneck can be more formally defined as the proportion of the total number of meaning-signal pairs seen by each learner agent. Note that this is a physical restriction on the *meanings that a learner encounters in the world*, and not a restriction on the signals that a speaker produces, or can accurately retrieve from memory. These distinctions will be important later.

In many ways this is an entirely reasonable way to model the transmission bottleneck. The fact that this training subset is always selected anew at random for every learner is good because it effectively captures the idea that there is natural variation in exposure to meaning-signal pairs between individuals: each agent gets a unique sample of the language, paralleling the fact that no two natural language learners ever receive identical exposures to language. The fact that there remains a large proportion of novel (i.e. unseen) meaning-signal pairs that the learner agent

---

[2] Although there are some models which claim to have infinite meaning spaces (e.g. Kirby, 2002a), they are infinite in virtue of containing recursive operations. At their core, they rely on a finite set of atomic meanings. In such models it perhaps makes more sense to quantify the bottleneck in terms of the number of positive training examples seen by the agent instead. As the experiments reported in this thesis only have finite meaning-spaces, we will refer to the size of the bottleneck as the proportion of meaning-signal pairs seen vs. unseen.

might have to convey is also quite realistic. Our immense productivity in language is one of the traits we most wish to understand, after all.

However, leaving aside the discussion of semantic bottlenecks for a moment, it turns out that there are other ways in which modellers can conceptualise a bottleneck on language transmission. Hurford (2002) surveyed the ILM literature and identified at least two additional types of bottleneck to have been explicitly implemented in models: production bottlenecks and intake bottlenecks (Table 4.1).

type	*description*
semantic	learners encounter just a subset of possible meanings during acquisition
production	speakers produce just a subset of possible utterances after acquisition
intake	only a subset of meaning-signal pairs are actually taken in and used in acquisition

Table 4.1: A summary of the definitions of three transmission bottlenecks discussed in Hurford (2002).

A **production bottleneck** appears as a result of choices made by the agent over which signal to produce in response to a given meaning. When speakers have acquired several different forms for a particular meaning, they must somehow decide which one to utter at any given moment, they cannot simply utter them all. An **intake bottleneck** on the other hand relates to the fact that not all of the meaning-signal pairs which are heard by a learner are actually used when it comes to the process of acquisition itself. Of the linguistic data to which a child is exposed, only a subset of it may trigger learning. We must therefore distinguish between input to a learner and what they take from that input (Hurford, 2002).

All of the models in Hurford's review seem to actually implement these different types of bottleneck in parallel, albeit often only implicitly. For example, in Kirby (2001) there was explicit mention of how the semantic bottleneck was implemented. A careful reading of the text reveals however that if agents had more than one rule in their grammar for conveying a particular meaning, only one would be selected at random for production, and that any forms which already had a meaning assigned to them were ignored by the induction algorithm if they were seen in a different meaning context -- in other words, production and intake bottlenecks had also been incorporated implicitly into the model.

In order to understand the individual contributions made by each type of transmission constraint, Hurford (2002) ran a series of simulations where bottlenecks were applied one at a time in a simple model of vocabulary evolution[3]. In these simulations, populations of agents learned names for a finite set of atomic meanings by observing other agents' naming behaviour. If agents were prompted to name an unfamiliar meaning, they could invent a random signal for it, but if the meaning had already been encountered, they could use a remembered name. He discovered that when there was a semantic bottleneck but no production bottleneck the number of synonyms in the lexicon tended to increase. This is because at every round there was a fairly high chance that a novel meaning would be encountered for which the agent had no signal. In this situation, the agent will have to invent a new signal, thus increasing the number of signals associated with that meaning in the population as a whole.

---

[3] In the end, Hurford did not run these simulations isolating intake bottlenecks, possibly because there are many different possible ways to conceive of constraints operating on language acquisition, all of which may have different signature effects when applied in isolation. However, with regards to the intake bottleneck implemented in Kirby (2001), he concluded that it worked in that instance to prevent homonymy from arising (Hurford, 2002). By ignoring forms they had already assigned a meaning to, agents avoided introducing ambiguity into their grammars. A different intake bottleneck (e.g. only signals with the letter 'a' will be acquired by the learning mechanism) would have very different effect (the number of signals with the letter 'a' in them would rapidly increase in the next generation).

In contrast, when there is only a production bottleneck, but no semantic bottleneck, the number of synonyms decreased until there was only one signal associated with every meaning in the population. This works irrespective of how the production bottleneck is implemented (e.g. whether the agent selects the form it has heard most frequently, the form it has heard most recently, or one chosen at random). Again, this is explained by the fact that there will inevitably be chance occasions where a particular form is not produced by a learner. Consequently, that form will not be present in that learner's output to the next generation, and over time, will disappear from the population altogether.

These findings present us with several intriguing ideas. First and foremost, it has been shown that different types of transmission bottleneck may have different effects on the structure of language. Secondly, these different constraints can work orthogonal to one another. If there is tension and trade-offs between them it means that we cannot take it for granted that their combined effects will simply be additive. If one constraint is encouraging synonymy, and the other attenuating it, we cannot predict in advance what the eventual outcome will be. This in turn means that we need to be even more certain that we design models with the right features. So what is the 'right' bottleneck to model? Is there a reason that past iterated learning models have tended to implicitly incorporate multiple bottlenecks? In order to answer these questions we need to look more closely at where these bottlenecks originate: in the transmission cycle itself.

4.1.3 Bottlenecks in the Transmission Cycle

In biology it has been noted that evolution can be seen as a series of transformations between different types of object - for instance, the transformation of genotypes into phenotypes (Lewontin, 1974). It is possible to view linguistic evolution as a series of transformations between different types of object as well (Kirby, 1999). In fact, I would like to argue that we can modify this idea slightly by thinking of these transformations as bottlenecks operating within different domains - the external "E-Language" domain where language is publicly represented, and the internal "I-Language" domain where language is privately represented in the minds of speakers and hearers.

The transformations that were originally identified in Kirby (1999) have been reproduced in Figure 4.1. I have relabelled this diagram with the different kinds of bottlenecks that Hurford (2002) identifies. Here, one complete cycle of transmission is shown, starting with a fully competent language user producing a set of utterances, and ending up with another language learner developing full linguistic competence.

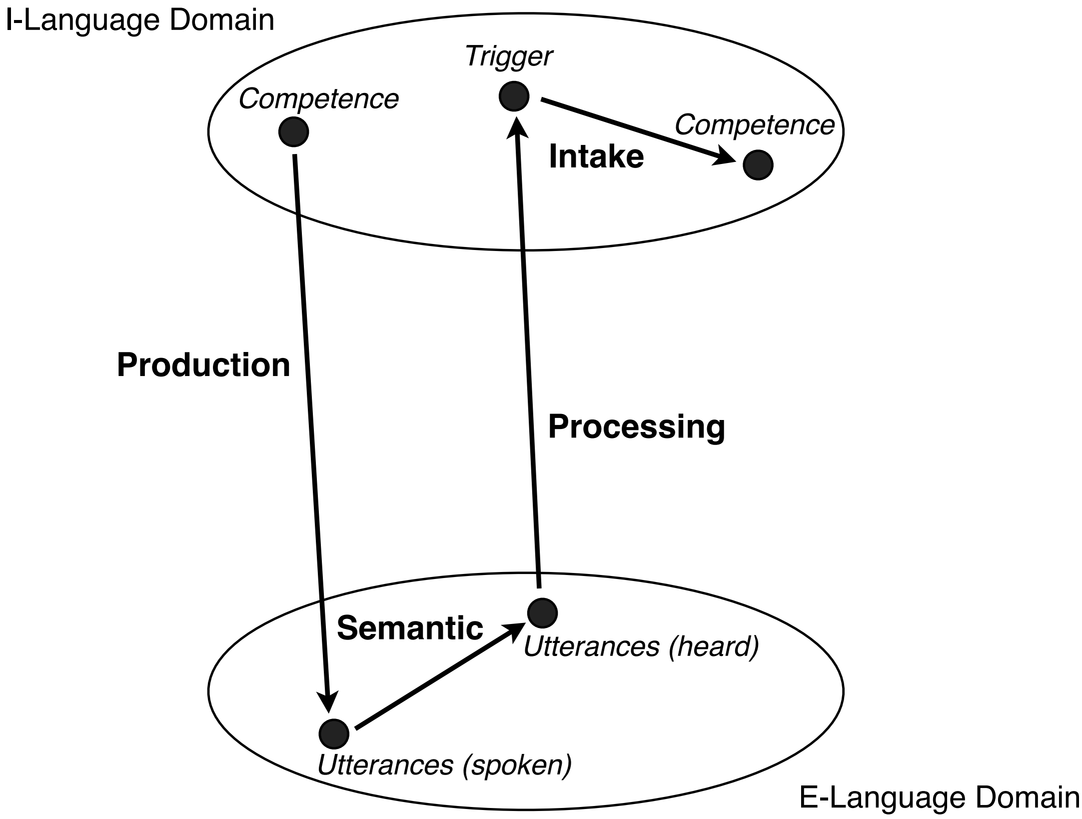


Figure 4.1: Reproduction of a diagram showing the transformations within and between the I-Language and E-Language domains of a speaker and a learner, taken from Kirby (1999). These transformations, which were originally just labelled T1, T2, T3 and T4, can actually be conceived of

as bottlenecks that appear through the cycle of transmission. In the I-Domain, language internalised in the mind of a speaker (competence) goes through a production bottleneck which determines which utterances get spoken and appear in the E-Domain. This set of spoken utterances then encounters a semantic bottleneck, which reduces this set to just the utterances heard by a learner. These heard utterances then go through a processing bottleneck, which filters out any utterances that are not memorable/salient. Out of this information which has been spoken, heard, and remembered, only a subset is hypothesised to actually be used in the learning process (to develop competence in a new individual).

Starting with the competent language user, we find that the production bottleneck mediates which utterances get transferred from the I-Language domain, to the E-Language domain. The process of production operates internally within a speaker. In contrast, as we have already mentioned, the semantic bottleneck is external to the agent. It determines which (out of a finite many) of the utterances that have been spoken by a language user, actually make it to the ears of a language learner. At this point we need to identify another kind of bottleneck, which I will call a **processing bottleneck.** This type of bottleneck covers the transformation from E-Language back to I-Language, whereby utterances that have actually made it to the ears of the learner get processed and parsed by the cognitive mechanisms that learner possesses[4]. Finally, a subset of whatever makes it through processing will eventually go on to trigger changes in the knowledge a learner has of language (intake). Both the processing and intake bottlenecks operate internally within the language learner. The result of all of this is the successful transmission of linguistic competence.

Of course, these bottlenecks are more than singular constraints. Instead, they are *types of constraints*. For instance, many factors influence which of the utterances that get spoken make it to the ears of the hearer; such as the presence or absence of noise, the structure of the environment (i.e. what there is to talk about), and even social dynamics between the speaker and hearer. Nevertheless, we can see all of these biases as a type of semantic bottleneck in virtue of which point it occurs at in the transmission cycle. Traditionally these constraints in transmission have all been considered to operate more or less simultaneously, as Figure. 4.2 shows.

---

[4] Kirby (1999) deals almost exclusively with this kind of bottleneck.

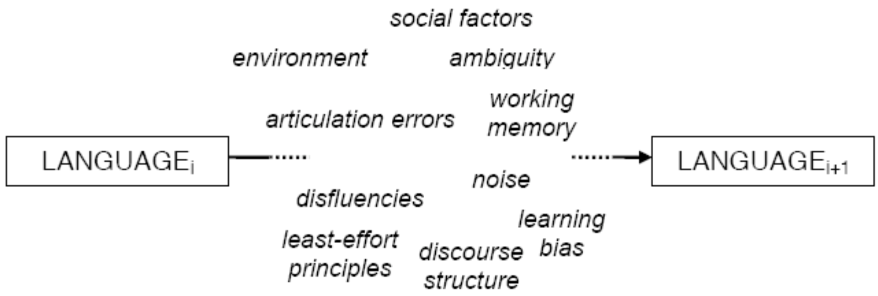


Figure 4.2: Traditional conceptualisation of the constraints operating during linguistic transmission. There are many different biases all acting at the moment of transmission. However, the cycle of transmission itself is composed of different moments. Understanding how these moments relate to one another can help us to design experiments to tease these factors apart. Taken from Kirby, Brighton & Smith (2004).

Although they undoubtedly do all make their influence felt at some point during transmission, we can perhaps be more specific about exactly when this occurs. This in turn allows us more control in designing experiments that investigate any of these specific topics. Whilst it is true that we cannot remove the internal bottlenecks from our participants[5], we can at least manipulate those bottlenecks which are external to the learner, and ask what influence they have on the cultural transmission of language. The rest of this chapter looks at this.

4.2 Experiment I: Generalisation to Novel Stimuli

Following on from our discussion of the role of bottlenecks in cultural transmission it should be clear that our goal here can not be to understand every bias affecting language. Instead, we can learn a valuable lesson from the computational simulation approach: by stripping away the complexities of the problem and

[5] Although we might be able to influence them in other ways, for instance, by forcing or ignoring certain types of production from our participants, or limiting their working memory by running distraction tasks during processing.

starting the modelling process off on a small-scale, we can come to understand what the essential components are and how much of the behaviour can be explained by the little things. We have already begun this process by creating an experimental framework capable of observing the evolution of simple forms and structured meanings in the laboratory, and breaking down the process of transmission into distinct stages that we can exert some experimental control over. Now it is time to see whether anything useful comes from this. We will start by attempting to replicate a common finding of computational models of iterated learning: that compositional structure arises when agents are forced to generalise to novel stimuli.

4.2.1 Method

An overview of the general methodology for the experiment can be found in §3.4.2. This section describes the particular design used to explore what effect being exposed to novel (i.e. unseen) stimuli has on the structure of the resulting languages. As always, the experiment is interested in how structure emerges from a state of non-structure, and it was important that participants were unaware that their data would be passed on to future learners.

Aims and Experimental Hypotheses

The basic aim of the first experiment was to try to replicate the computational findings concerning the semantic bottleneck in a small population of human learners. To reiterate, ILMs with linear transmission chains and a semantic bottleneck in place (e.g. Kirby, 2000) tend to result in systems which are both highly learnable, and highly structured. In particular, it has been found that compositional structure emerges from an initially holistic language. That is, agents converged on a solution to express complex meanings by using signals where the meaning of the whole was derived from the meanings of signal-parts and the way they were put together.

Following from this, our expectations were as follows. If human learners were actually doing the same thing as simulated agents, we should first expect that the languages being transmitted between human agents should become easier to learn toward the end of the experiment: that is, transmission error scores between learners should decrease as the number of generations increase. The second result we expect is that this decrease in transmission error should correlate with an increase in the amount of structure in the languages. Finally, we might also expect to see compositional structure emerging. This leads to the following hypotheses:

1. **The Learnability Hypothesis**: Languages will become easier to learn as a result of iterated learning.
2. **The Structure-Increase Hypothesis:** Languages will become more structured as a result of iterated learning.
3. **The Compositionality Hypothesis:** Pressure to generalise to novel stimuli will result in languages evolving to become compositional.

Experimental Design

In order to test these hypotheses, a series of four transmission chains, each consisting of ten 'generations' of learners, were run[6]. Each chain was initialised with a different randomly generated initial language, and all used the same structured 3×3×3 meaning-space, as described in §3.4.2. As we wanted to try to replicate the computer simulations, a 50% semantic bottleneck was also implemented. Given that there is an odd number of meanings, this meant that each participant was trained on exactly 14 out of the 27 items. These training items varied between generations, and were selected at random anew during the off-line transmission phase of the previous generation.

---

[6] One of the chains in this experiment (C) was actually obtained during my MSc project, where I piloted this framework (Cornish, 2006).

Three rounds of training were given, with each of the 14 seen items being presented twice within each block. In between training rounds, there was a test phase. The first two tests were short -- containing only 7 seen and 7 unseen items. The final test consisted of every single meaning. Training data for the next generation was drawn exclusively from this set of final responses. Figure 3.1 in §3.4.2 shows a graphical representation of this training-test schedule. The experiment itself was run using E-Prime, and the statistics were analysed using *R.*

Participants

In total, 40 participants were recruited, primarily via an advertisement placed in the University of Edinburgh's student employment service (age: M = 22.25, $S.D$ = 2.43). There were 25 female participants, and 15 male participants, and each was assigned to a chain and a generation at random. Participants were not required to be monolingual English speakers, but were required to be fluent in English. In order to be eligible to take part, volunteers had to have normal or correct-to-normal vision, not be dyslexic, not have already taken part in an 'alien language learning' experiment before, and also not have formally studied linguistics. This latter requirement was added following piloting of the software used for presenting the experiment, which revealed that those with an extensive background in the formal analysis of linguistic systems tended to approach the task in a highly analytical way and were more likely to have come across the iterated learning models or have some idea of the research interests of the experimenter.

Finally, participants were compensated with £5 for their time and travel costs. The study conformed to the ethics guidelines set by the University of Edinburgh's College of Humanities and Social Science, and participants were fully briefed before taking part of their rights of withdrawal and anonymity.

Procedure

Instructions were given both in writing (see Appendix A) and verbally. Participants were told that they would be shown a series of images and the way in which an alien would describe those images, and that after some time, they would be tested to see what they had learned. However, they were unaware that their output data would become training data for future learners. They were encouraged to always give a response, even if they were unsure “in order to maintain good relations with the aliens”. They were informed that there would be three rounds of training and testing, and that although the training items were automatically presented by the computer, they could pace themselves throughout all the testing phases.

At the end of the language learning task, participants were given a short questionnaire to fill in detailing what they thought the experiment was about, how they approached the task of learning the language, how they thought the language worked, and at what point (if any) they became aware that they were being tested on items that they had not been trained on. Finally, once this data had been collected, participants were fully debriefed about the experimental aims of the study. All in all, each experiment lasted no more than 35 minutes, including the questionnaire and debrief. The experiment was run on E-prime, and the results were analysed using *R*.

4.2.2 Results of Experiment I

Structure and Learnability Increase

In order to address our first two hypotheses, the structure and error scores were calculated for every generation of each chain and are plotted in Figure 4.3 below (see §3.4.3 for how these measures are calculated). The graph on the left shows the *z*-scores calculated from running the Mantel Test with 10,000 randomisations. As we can see, structure increases in each chain. In particular, all languages were significantly structured after generation 6 (the dotted line represents the 95% confidence interval - subsequently, any point above that line is significant to the $p<0.05$ level or greater). The graph on the right of Fig. 4.3 shows the learnability of the languages in each chain in terms of the average nLD error score between adjacent generations. This shows a clear decrease in transmission error over time, with half of the chains becoming stable in the final generations.

In order to determine whether there was a significant increase in structure and decrease in error over the course of the whole experiment, one-tailed paired t-tests were run on the beginning and ends of the chains[7]. From the results of these it was possible to confirm both predictions: the languages are adapting to become significantly more structured over time (as shown by a mean increase in structure of 4.763, $t(3) = 3.4296$, $P < 0.02$) and significantly more learnable over time (as shown by a mean decrease in transmission error between first and final generations of 0.638, $t(3) = 8.656$; $P < 0.002$).

[7] Two data points are missing from Chain A in Fig 4.3.left. This is because at this point the language had only two words in it - one for 'blue spiralling square' and another for everything else. In this situation, it does not matter which meaning the odd-signal-out is associated with: in terms of the degree of structure this kind of system has, the same value would be returned if the odd-signal-out was paired with *any* of the meanings. Following this intuition, this means that in this situation our measure of structure is undefined. The Monte Carlo sample (which we use to produce our *z*-score) contains no variation on account of every possible ordering being equally likely. The t-test for structure was therefore run using the score for generation 8 instead of 10 for that chain.

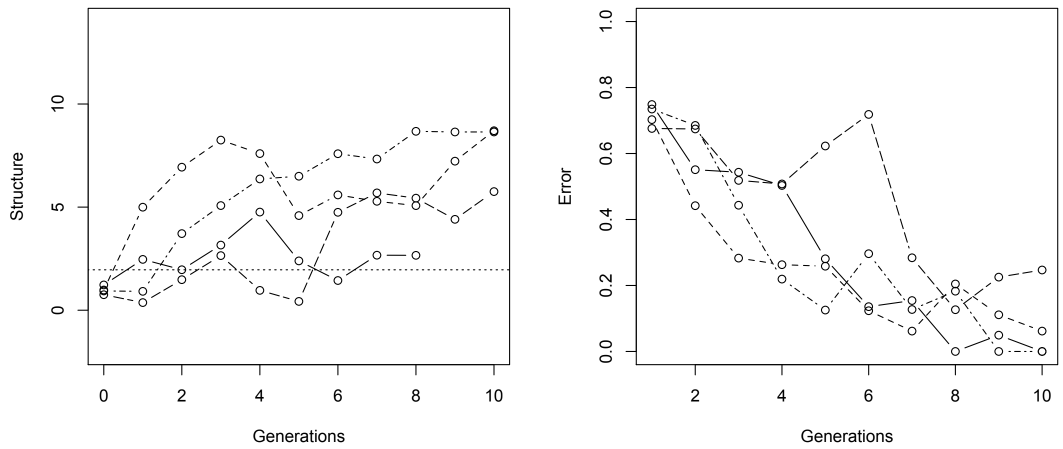


Figure 4.3. Graph showing the structure (left) and normalised error (right) scores by generation, of four transmission chains where a 50% data bottleneck was present. These results indicate that languages are becoming significantly more structured and easier to learn over time. Points above the dotted line (left) represent significant structural regularities between meaning-signal mappings. This graph has been remade from Kirby, Cornish & Smith (2008) with 10,000 Monte-Carlo randomisations instead of 1,000.

This begs the question of exactly *how* the languages are changing to do that. In computer simulations, the presence of a semantic bottleneck encourages systems to arise that are compositional. Have the languages in the experiments evolved to become compositional as well?

Reduction of Signal Types

If we examine the total number of distinct signals used at each generation, we find that this cannot be the case. Table 4.2 shows that this number decreases both rapidly and dramatically across all chains. If the languages were perfectly compositional, there would be 27 distinct signals for each of the 27 distinct meanings. However, although it is tempting to conclude that this decrease in the number of strings can explain the reduction in error seen earlier -- as having fewer strings to learn makes the process of recall easier -- the presence of the semantic bottleneck means this cannot be the case. Remember that participants are only being trained on half of the meaning-signal pairs. In order to achieve perfect transmission (which some chains do), this entails that there is intergenerational agreement on the signals to be used on unseen meanings. Even if we only have a handful of distinct signals to remember, it is not obvious how this alone would help bypass the constraints imposed by our transmission bottleneck.

generation	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
chain A	27	17	9	6	5	4	4	2	2	2	2
chain B	27	17	15	8	7	6	6	6	5	5	4
chain C	27	24	8	6	6	5	6	5	5	5	5
chain D	27	23	9	10	9	11	7	5	5	4	4

Table 4.2: Table showing the number of distinct signal strings found in the languages of each generation for four chains. This number decreases very quickly over time, resulting in just a handful of unique signals at the end of the experiment.

Systematic Underspecification: An Adaptation

In order to understand what is going on we need to move away from quantitative analyses of the languages, and start examining them qualitatively. Table 4.3 shows the final language from one of the chains resulting in a stable language. Each cell in the table shows the signal used for each individual meaning, with motion and shape features indicated in the corresponding row names, and the colour feature being represented by column. As we would expect given Table 4.2, instead of a one-to-one mapping between meanings and signals, we find ambiguity. However, the meanings are not just underspecified by the signals, they are *systematically underspecified* by them. In a way, systematic underspecification is a type of categorisation. Although there are just five signals, there is a regular pattern in the way those five signals get assigned to the meanings. In this particular case all spiralling objects are called ‘poi’, all horizontally moving objects are called ‘tuge’, and there is a three way distinction between bouncing objects based on the shape of that object: squares are 'tupim', circles are 'miniku', and triangles are 'tupin'.

Clearly the fact that this system persists completely unchanged for the final three generations indicates that it is well adapted to the problem of being faithfully transmitted. Rather than treating the ambiguity as in instance of homonymy or synonymy though, it is possible that the meaning-space itself is changing along with the language. Support for this theory comes directly from the post-test reports of learners exposed to this particular system. According to the participant at generation 9 of this chain: "the aliens don't seem to care about colour". For this learner, although he could clearly perceive that the meanings varied along three-dimensions, the signals themselves forced him to reinterpret that assumption. By rationalising that the aliens were colour-blind, he realised that there were in fact, only two dimensions to keep track of.

tuge	tuge	tuge	□
tuge	tuge	tuge	○
tuge	tuge	tuge	△
tupim	tupim	tupim	□
miniku	miniku	miniku	○
tupin	tupin	tupin	△
poi	poi	poi	□
poi	poi	poi	○
poi	poi	poi	△

Table 4.3: A table showing the final language from a stable system in Experiment I (Chain C). The meanings in this language are systematically underspecified by the signals. This system easily survives the transmission bottleneck by effectively reducing the number of meanings to just five: things that move horizontally; things that spiral; bouncing squares; bouncing circles; and bouncing triangles. Given that the bottleneck allows 14 items through, the odds of at least one item from each of

the five emergent categories surviving are high. This table has been redrawn from Kirby, Cornish & Smith (2008) with permission.

So how does this actually make the language easier to learn? By dropping a meaning feature like colour (and in the case of spiralling and horizontally moving objects, shape also) the system has not only decreased the number of salient features to be differentiated by name, it has also effectively increased the number of possible tokens of each 'type' or category of meaning. To explain, there is only one token of a horizontal black square, but there are three tokens of a horizontal square, and nine tokens of something horizontal. By increasing the number of tokens for a given meaning, you increase the frequency and likelihood of it passing through the semantic bottleneck to be reproduced by the next generation. Systematic underspecification therefore appears to be a powerful adaptation, perhaps explaining why it appears to some extent in all four chains. We will return to this notion in more detail later in the discussion, but for now, we can ask ourselves how it is exactly that systems like this come to arise.

The Evolution of Signal Forms

One of the exciting things about iterated language learning experiments is that we are able to live the diachronic linguist's fantasy: we have a continuous and complete record of the utterance acquisition and production history of every speaker in a language, and we can use this to find the early origins of synchronic features of the language. For instance, if we examine the history of the language shown in Table 4.3 we can trace the changes each and every signal underwent over time. If we pick one of the signals in the final generation, for instance, 'miniku' (meaning 'bouncing circles') we can follow its ancestry right back to a variant in the original input: 'miniki', meaning a horizontal blue square. This form was altered to 'miniku' by the very first learner, again to refer to a horizontal blue square. It wasn't until generation 4 that this signal became associated exclusively with bouncing objects, but then it was mostly used for bouncing triangles. By generation 7 it was being used for nearly all bouncing items, but at generation 8 it appeared in its current role, referring exclusively to bouncing circles (see Appendix B1).

These historical changes over time can be more succinctly represented visually in a coalescent tree (Cornish, Tamariz & Kirby, 2009). These trees are used extensively in evolutionary biology to show relationships of descent amongst phylogenies (Barton, 2007; Hein, Scherup & Wiuf, 2005). One potential issue is how we determine relationships of descent in this instance. The 'miniku' example was fairly trivial to analyse, as the signal itself underwent almost no changes and could be traced back to a variant in the initial signal-set by virtue of a common meaning. However, not every signal will be this free of noise. We have already seen that the mappings themselves, between signals and meanings, are adapting to transmission constraints and are therefore highly changeable. Given that our goal is only to trace the evolution of signal *forms* over time, and not the mappings, we need to factor out the mappings from our analysis entirely[8].

How then are we to proceed? We need to start by making some simplifying assumptions. The first is that a learner's representation of a particular form is potentially influenced by *any* of the forms that they have seen during training, and not just the target one. The second, is that in cases where we see exact replication of a signal, we can confidently assume a relationship of descent exists, and in cases where we see only similarity with other signals, we can only assume a possible relationship of descent. Although we can operationalise this similarity algorithmically (for instance, by only classifying signals that are above a certain nLD threshold as being in a possible relationship of descent), in this experiment the signal-sets are actually small enough to analyse by hand[9]. Figure 4.4 shows a

---

[8] If we do not then we are simply reproducing the data in the way it is shown in Appendix B1, which tells us which signals were used for which meaning and how that changes over time, but tells us nothing about how new signals are innovated or altered.

[9] Multiple coders can be used, and measures of inter-rater reliability taken to control for any effects of coding bias. Given the fact that we are making similarity judgements with a maximum of just 14 strings, just one coder was used on this data.

coalescent tree generated for chain A from the first experiment. Undisputed relationships of descent are shown by solid lines, whereas possible relationships are indicated by dotted lines. Obviously only seen items (shown in bold) could influence the language of future generations[10].

From the tree we can see that there is initially a lot of variation and innovation going on, with very little faithful transmission, as evidenced by the number of dotted lines indicating possible relationships of descent. At this early stage in the history of the language, the transmission process seems to principally involve the generation of new signals out of recombinations of signal parts. Rather than witnessing the replication of whole signals, we see replication of bigrams and larger n-grams in new configurations. For example, in the first generation we find the signals 'lepa' and 'pali' arising. These signals were both present as substrings within a seen signal in the initial language: 'lepali'. In addition to finding innovations resulting from the loss of signal elements, we also see innovations arising from blends. For instance, the appearance of 'nepi' in the third generation appears to be the result of a mixture of 'nemi' and 'nepa'.

After this brief flurry of innovation, we quickly see a 'core' set of signals developing. This occurs from around generation three onwards. One thing that is particularly noticeable in this tree is the strong effect of 'frozen accidents' (Gell-Mann, 1994). A frozen accident is essentially a chance event which has far-reaching consequences for the future. We can see this when, for instance, the random selection of seen items causes signals to be completely lost from the system. Importantly, this occurs irrespective of whether those signals appeared particularly frequently in the populations before they were selected.

---

[10] Those with keen eye-sight will notice a curious example in Fig. 4.4 whereby an unseen item - 'nemi' - was nevertheless perfectly reproduced by the second learner. This is not a graphing error, but is a phenomenon that occurs several times in the data. Although it is unusual to see it in the very first generation, there is nothing particularly miraculous about its appearance here: two other seen signals just happened to be very similar to it ('nemine' and 'lemi').

Why is this important? Well, in this instance we can think of numerosity as a proxy for the fitness of each signal. Their suitability as signals has been tested by past learners, and their frequency has actually increased. If two signals are equally fit (frequent), and one gets selected and ends up heavily influencing the system in the future whilst the other does not, then we must conclude that this is an instance of a historical accident. It was not the case that one signal was better suited to be passed on, it was just a chance event that led one to propagate, and the other to become extinct. On the other hand, if two signals were not equally fit, and the fitter one went on to affect the future language, this would instead be an instance of cultural selection. We see examples of both processes at work here: for instance, the loss of 'lepa' and the survival of 'nemene' in generation 2 appear to be the product of chance; whereas the relative success of 'maho' over 'mapo' in generation 1 might be better construed as cultural selection. Crucially, however it is the random application of the bottleneck that would seem to be most responsible for the steady loss of variation in signals. We will return to this idea in §4.4.

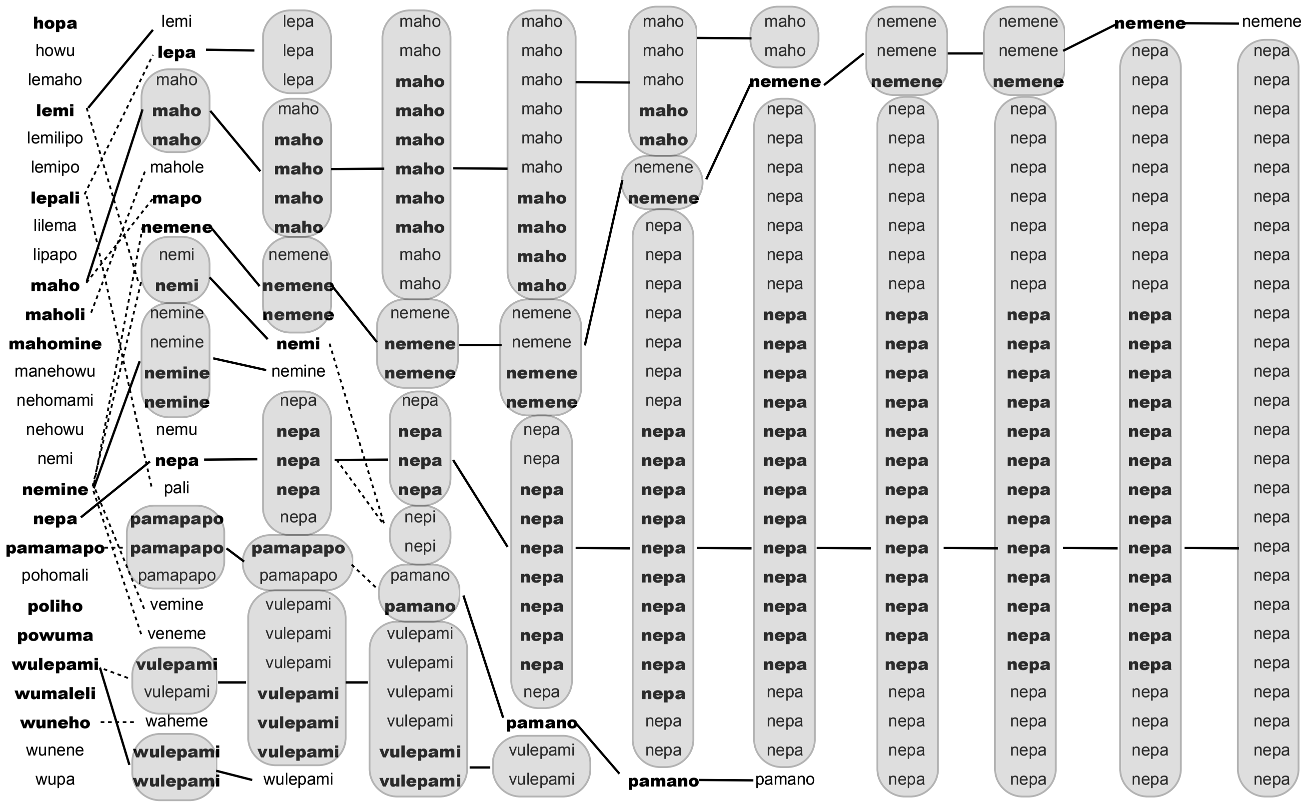


Fig. 4.4: A coalescent tree showing relationships (solid lines) and possible relationships (dotted lines) of descent between signals in chain A of Experiment I. Signals which were seen by the next generation are indicated in bold, and repeated signals are grouped together for clarity. This tree demonstrates how the bottleneck encourages signal variation to diminish over time.

4.2.3. Summary

In summary, we can accept both the learnability and the structure-increase hypotheses set forward earlier. We found that the transmission error between languages significantly decreased over time, and that this coincided with a significant increase in the amount of structure found in each language over time. However, both quantitative and qualitative examinations of the languages themselves revealed that unlike in computer simulations of iterated language learning, the systems did not evolve to become compositional. Instead, individual signals increased their chances of surviving the bottleneck by increasing their frequency in the languages. This came at the expense of expressing all 27 meanings, which meant that the meanings became underspecified by the signals. However, this underspecification was not indiscriminate: it was argued that a systematic relationship between signals and meanings still evolved, it just involved a reclassification of the meaning-space to describe fewer dimensions. We must therefore reject the compositionality hypothesis, and conclude that there is something slightly different going on in the human iterated language learning study, compared to the computational versions.

4.3 Experiment II: The Effect of Imperfect Learning

The first experiment has shown us that when there is a semantic bottleneck in place it creates a pressure for greater generalisation. Under such circumstances, systems will actually adapt to overcome the restrictions imposed by this bottleneck, and reorganise themselves in a systematic way. The fact that participants are being forced to describe novel stimuli during their final test however, makes it impossible for them to fully succeed at the task without some sort of structural relationship existing between meanings and signals. Even though the participants are not aware of it until the last minute (if at all), this simple experimental manipulation has made a memorisation strategy completely useless. In some sense this is unfair: participants are being misled as to the parameters of the task they have been given, and the design of the experiment is such that some sort of change in the language is inevitable from the outset. Put another way, it is all well and good to show that participants generalise when faced with labelling novel stimuli, but can we be sure that this situation arises naturally without some reliable intervention?

In order to address this concern, a second experiment was run. This time, instead of only training participants on half of the language, they were exposed to all of it. This means that, in principle, it is now possible for a language to be memorised and transmitted holistically right from the beginning of the experiment. This has been tried before, in computational simulations of iterated learning. For instance, Smith (2003) clearly demonstrates that if an ILM is run with no semantic bottleneck in place, no cultural evolution occurs at all -- the randomly generated idiosyncratic signals created in the first generation are maintained throughout. This is because in this model, agents are perfect learners. They are capable of memorising even large numbers of meaning-signal pairs flawlessly. Given the fact that human memory is not that reliable, it is worth investigating whether this 'memory bottleneck' could play any significant role in the emergence of structure[11].

4.3.1 Method

Aims and Experimental Hypotheses

Although our research question is different (instead of questioning whether the semantic bottleneck is capable of explaining the emergence of structure in language, we are now questioning whether natural human memory limitations can function in

---

[11] To use to terminology outlined in §4.1.3, constraints on our memory are clearly a kind of processing bottleneck, affecting how much of the information that survives the semantic and perceptual bottlenecks actually makes it into our heads.

the same way) our experimental hypotheses remain basically the same, with one exception. Given that in the last experiment, the languages evolved to be systematically underspecified instead of compositional, we will replace the compositionality hypothesis with one about systematic underspecification.

1. **The Learnability Hypothesis**: Languages will become easier to learn as a result of iterated learning.
2. **The Structure-Increase Hypothesis:** Languages will become more structured as a result of iterated learning.
3. **The Systematic-Underspecification Hypothesis:** Pressure arising from natural memory constraints will result in languages evolving to become underspecified in a systematic way.

Throughout, we are looking for any similarities or differences in the kind of languages that emerge in the first and second experiments. Because the only experimental difference between them is the presence or absence of the semantic bottleneck, we can consider them to be two contrasting conditions. Occasionally it will be more convenient to refer to the two studies by condition (i.e semantic bottleneck (SB) or no semantic bottleneck (nSB)) instead of by name.

Experimental design

Once again four transmission chains were set-up with new sets of initial languages, randomly generated and paired with the structured 3×3×3 meaning-space. These chains were each run for ten generations of learners. Unlike the first experiment however, each participant was trained on all 27 meaning-signal pairs. Training and testing phases were held proportional to those in Experiment I; although there were more training items in total, each meaning-signal pair was seen the same number of times (i.e. six times - twice in each of three rounds) and for the same duration (six seconds total) as before. Following the final test, the entire set of meaning-signal pairs was transmitted to the next participant. Again, this experiment was carried out using E-Prime, and the results were analysed using *R*.

Participants

A total of 40 new participants (age: M = 21.075, $S.D$ = 2.63) were recruited via another advertisement in the University of Edinburgh's student employment service. From this 40, there were 24 female and 16 male participants. Eligibility restrictions were the same: normal or corrected-to-normal vision, no dyslexia, participating in an 'alien language learning' experiment for the first time, fluency in English, and not a linguistics student. These participants were paid £5 for their time and travel costs, were randomly assigned to one of the four language chains, and were given the exact same instructions as the previous participants. All participants were fully briefed before taking part of their rights of withdrawal and anonymity. This study fully conformed to the ethics guidelines set by the University of Edinburgh's College of Humanities and Social Science.

Procedure

The experimental procedure was identical to that described in §4.2.1, with the only difference being that there was no semantic bottleneck in place. As the participants were being trained on more signal-meaning pairs however, Experiment II lasted a bit longer: including time for filling in the questionnaires and debriefing, participants spent around 50 minutes on the task.

4.3.2 Results of Experiment II

Structure and Learnability Increase

In spite of the changes we have made to the experiment design, our hypotheses concerning the languages becoming more structured and easier to learn remain the same. To determine whether they hold, the structure and error scores were calculated for each language chain, and are shown in Figure 4.5 below. Overall, we find a similar result to the first experiment, with all chains showing statistically significant increases in structure (mean increase of 7.39; t(3)=9.08, $P<0.001$) and decreases in error (mean decrease of 4.45; t(3) = 4.628, $P<0.005$) between the first and last generations, as shown by one-way paired t-tests. However, the systems that emerged do appear to be less stable than the ones from the previous experiment. Only one of the four languages was perfectly transmitted, and only for one generation.

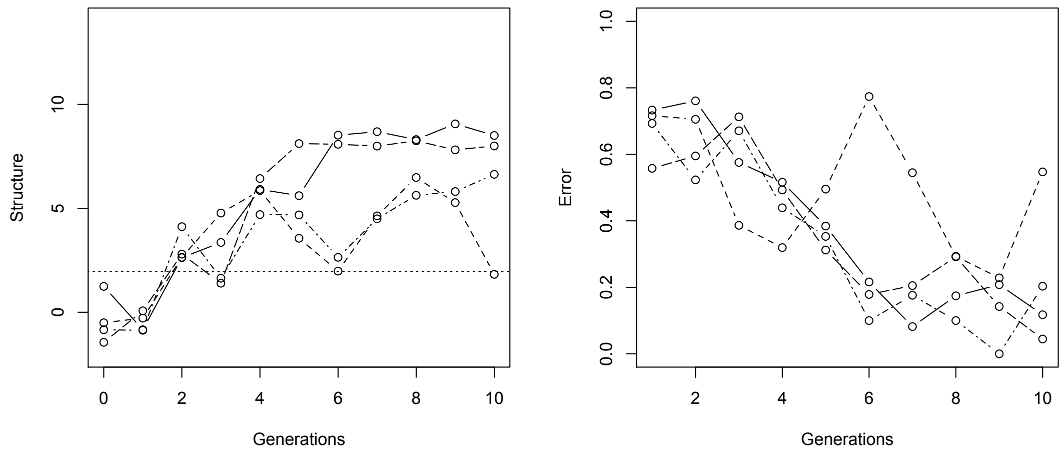


Figure 4.5: Graph showing the learnability and structure scores over generations in Experiment II. The graph on the left shows that structure increases steadily over time (points above the dotted line represent z-scores significantly different from what we would expect by chance). The graph on the right shows transmission error (measured as nLD) decreases over time. Both graphs have been remade from Cornish (2010), using 10,000 Monte-Carlo randomisations instead of 1,000.

Finding less stability in the systems is in some ways counter-intuitive, as we might expect that once some structure had emerged in the languages, having full access to the data should actually facilitate acquisition. Afterall, it should be easier to detect structural regularities when provided with more evidence than it is with less. Of course, a general point can be made here concerning the fact that we only have four chains to base our observations upon. We cannot rule out the possibility that we have just been unlucky not to observe more stability in this particular instance. However, it could also be that there are features within the languages themselves which actively discourages stability from emerging. It is to this question we now turn.

No Change in Level of Difficulty of Task

Perhaps the reason for instability is simply due to the fact that we are presenting the participants with a harder task. It is possible that even though we have kept the amount of training per item constant, by giving participants access to more meanings to learn from we have increased the cognitive load placed upon them. If this prevents useful structures emerging in the first place, then having extra training data could actually reinforce the idea that the system contains no order and slow down the whole process. In order to test this we can compare the learnability scores for the first generation of learners in the current experiment (shown in Fig.4.5.right) to the scores found for the first generation in the previous experiment (shown in Fig. 4.3.right). If participants are finding the task harder in the nSB condition, we would expect to see that reflected in their performance on the random initial languages.

For clarity, the error-scores of the first generation for the four chains in each condition are reproduced in Table 4.4 below. From this table we can see that there is not much difference between the performance of the first learner in either condition. Although on average we find recall to be slightly better in the nSB condition (mean error of 0.715 with the semantic bottleneck, and just 0.675 without), this difference was not found to be statistically significant ($t = 0.9389$, $df = 3.974$, p-value $= 0.4012$). This means that exposure to a greater number of meaning-signal pairs is not making the task more difficult for participants.

Interestingly, the statistical test also indicates that seeing the entire language is not making the task of learning the system any easier either. This is surprising, as when the bottleneck is in place the performance of the first participant is rigidly capped. In order to get error lower than the 0.5 mark, the participant would have to have access to the unseen meaning-signal pairs, which of course, they do not. Participants in the second experiment on the other hand, have no such restriction. There is nothing in the design of the experiment preventing them from doing much better than 0.5. Obviously whilst there are some talented individuals in the world who are capable of associating 27 randomly constructed signals with 27 unfamiliar meanings in the 16 minutes and 20 seconds of training time given, most individuals are not. It seems that no matter how many examples participants are given, on average only approximately 30% of it will be correctly recalled[12].

Experiment I (SB)	*Experiment II (nSB)*
0.7484568	0.7330247
0.7026749	0.7161817
0.7347884	0.6929012
0.6759259	0.5578483

Table 4.4: Error-scores of the first learners in the semantic bottleneck (experiment 1, left) and no semantic bottleneck (experiment 2, right) conditions. The differences in these scores are not statistically significant, indicating that the removal of the bottleneck has not made the task of learning the language any easier or harder.

Reduction of Signal Types

In order to determine whether languages were becoming underspecified as in Experiment I, we can examine the number of distinct strings again. As Table 4.5

[12] Remember that the measure of error is based on the average recall of the whole language. Recalling 30% does not necessarily entail getting 30% of the signals correct. It could just as easily mean getting 30% of *each* signal correct, or some combination of perfect recall of a few signals and low levels of recall of parts of others.

shows, the number of distinct signal types once again decreased, but not quite to the same extent as previously found. In order to judge this, I ran another set of unpaired t-tests on these numbers, and those obtained from the first experiment (shown in Table 4.2) by generation. Although generations 1-8 showed no significant difference, in the final two generations the nSB condition had a statistically significant increase in the number of distinct words produced as compared to the SB condition: generation 9 ($t(3) = 2.4804$, $p=0.05$) and 10 ($t(3) = 3.1623$, $p=0.02$) have retained more signal types overall than what we witnessed in the previous experiment.

generation	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
chain A	27	23	20	14	13	13	10	9	10	7	7
chain B	27	21	14	10	8	8	11	9	9	9	6
chain C	27	20	15	14	8	6	6	5	6	6	7
chain D	27	25	10	8	5	5	5	5	5	5	5

Table 4.5: The number of distinct signal-types by generation for all chains in Experiment II. This table shows that there is a steady reduction in the total number of signals found in each language over time. However, there was a significant increase in the number of distinct signals in the final two generations when compared to the results of Experiment I.

Overall we can confirm our two basic hypotheses about learning and structure-increase, and so far have seen that removing the semantic bottleneck appears to be increasing the amount of variation in the systems, albeit at the expense of stability. Now we will look to see what a qualitative analysis of the languages in the nSB condition reveals.

Systematic Underspecification, Irregularity and Internal Structure

Table 4.6 shows the language which got the highest overall structure score, at generation 9 in chain A. From this, we can see that once again, a form of systematic underspecification has arisen. In the previous study, the appearance of this phenomenon was explained in terms of the bottleneck on transmission encouraging a reduction in the size and structure of the meaning-space: this in turn increased the number of tokens of each meaning-type, which increased the chance of at least one token making it through the bottleneck. This is enough to allow the system to be reliably inferred by the next learner. Here it appears that the 'memory bottleneck' is doing something similar. Just as in the example shown in §4.2.3, we find that the colour dimension has been lost[13]. However, there are a number of interesting differences here as well.

The first thing to notice here is that the signal form used for bouncing objects contains a number of irregular variants -- specifically, 'mucapo' and 'nukapo'. Although these signals all share roughly the same structure, they display variation in the first and second consonants. We do not find these kinds of irregulars in the semantic bottleneck condition (see raw data in Appendix B1). Furthermore, if we examine the signals associated with bouncing objects in the generations preceding this one, we find that this is not a one-off occurrence. The variation has actually persisted for some time (Table 4.7). Even as far back as generation 4 we find the same basic pattern. By looking at Table 4.7 we can see that these irregulars are not stably associated with any specific meanings (i.e. colour or shape). To borrow an analogy from phonology, it appears that they are in free variation, and although many of these alternative variants eventually disappear over time, they do so only gradually.

[13] It is worth pointing out that there is no evidence in my data to suggest that there is something special about colour which makes it more likely to be the dimension that gets ignored. Of the eight chains described so far, four could be described as motion:shape systems (where motion was most consistently encoded, and shape a secondary characteristic), two were colour:motion systems, one was a motion:colour system, and one was undetermined. The undetermined chain was the one shown in Fig. 4.3 which had just one word for 'blue spiralling square' and another for everything else.

hapo	hapo	hapo	□
hapo	hapo	hapo	○
hapo	hapo	hapo	△
nucapo	mucapo	mucapo	□
nucapo	nucapo	nucapo	○
nucapo	nukapo	mucapo	△
nuakini	nuakini	nuakini	□
wagini	wagini	wagini	○
waginini	waginini	waginini	△

Table 4.6: The language with the highest structure score in Experiment II (Chain A, generation 9). Although this language is systematically underspecified, there are some signs of internal structure to the way signals are constructed. For instance, all bouncing objects share the suffix '-ini', whereas all objects moving towards the right have the suffix '-apo'. There are also irregulars present (e.g. 'nukapo', 'mucapo') which we did not see evidence for previously. This table has been redrawn from Cornish (2010).

generation / meanings	4	5	6	7	8	9
black circle	muhapo	magini	nucapo	nucapo	nukapo	nucapo
blue circle	nucapo	mukapo	mukapo	mucapo	mukapo	nucapo
red circle	muhapo	mucapo	mucapo	mucapo	mucapo	nucapo
black square	mutapo	nucapo	nucapo	nucapo	nucapo	nucapo
blue square	mukapo	mugini	mukapo	mucapo	nukapo	mucapo
red square	muckapo	mucapo	mukapo	mukapo	nucapo	mucapo
black triangle	mugeni	mugenini	nucapo	nucapo	nukapo	nucapo
blue triangle	mukapo	mucapo	mukapo	mukapo	nucapo	nukapo
red triangle	muhapo	nucapo	mucapo	mukapo	nucapo	mucapo

Table 4.7: Variation in signals associated with bouncing objects in chain A. Although there is a common underlying pattern here (a nasal consonant, followed by 'u', followed by a non-nasal consonant, followed (mostly) by the sequence 'apo'), there appears to be nothing conditioning exactly which variant gets used. Irregulars like these can only survive in the language when there is no semantic bottleneck in place.

The presence of irregular variants can potentially explain why there is less stability here than when there is a semantic bottleneck in place: having full access to the data allows these irregulars to survive, and complicates the learning process by making it necessary to memorise exceptions on a case by case basis. Although broad categories of signal-types may be stable over time as Table 4.7 shows, our measures of error are highly sensitive to even slight differences in form, obscuring the fact that there is actually quite a lot of continuity between generations (Cornish, 2010).

The second thing to notice about the language shown in Table 4.6 is that it appears there is a degree of internal structure to the signals used. Spiralling objects all contain the suffix '-ini', whereas objects moving to the right all end in '-apo'. There also appears to be some local regularity associated with the shape of bouncing items: the prefix 'nuak-' refers to squares, 'wag-' refers to circles, and 'wagig-' refers to triangles. This is exciting, as it looks as though we are starting to see evidence for some kind of compositionality, albeit as part of a language that is very underspecified. Unfortunately none of the participants from this chain reported being aware of any kind of prefixes or suffixes, and indeed, the language loses this clear structure by the next generation.

Examining the other three chains in this condition (see Appendix B2) does not reveal any other cases of internal structure in regular alignment with particular meaning aspects, but all of the chains' signals do show signs of being composed of subparts that get reused. For instance in the final generation of chain C we find the majority of the words have a common ending, '-laki', and that the single exception to that pattern, 'mano', is itself repeated in another signal, 'manolaki'. We should be wary of getting too excited by this finding however, as the same thing occurs (although arguably to a lesser extent) in the languages created in the first experiment (Appendix B1)[14].

The Evolution of Signals

Two factors seem to differentiate the languages emerging in this condition: the signals contain more variation, and in addition to becoming more learnable via the underspecification route, at least one of the chains contains signals that have internal structure which unambiguously maps onto specific meaning aspects (like motion). We can examine how these features arise historically in the chains by creating another coalescent tree like in Figure 4.3. As both of these features are only found in one chain (A), we will continue our focused exploration of it here by examining the evolution of its signal forms. This coalescent tree is shown in Figure 4.5.

[14] Recall 'tupim' and 'tupin' in the language shown in Table 4.2. Also, the final few generations of chain B in Appendix B1 shows considerable reuse of signal parts: 'ninalehe', 'lehe', 'nina', 'wina', and 'winako' being prime examples.

One of the first things to notice if we compare the two coalescent trees is that although we appear to have more variation and innovation in the nSB condition overall, we again see that the trend in both is for more of this early on, rather than later. This is in spite of the fact that earlier we only found statistically significant differences in the number of distinct signals in the final two generations. How can we explain this? It seems to boil down to a simple case of numbers. The major difference between the two conditions is that when we have a semantic bottleneck in place, even easy to learn signal types are at risk of being removed from the language if they are infrequent.

The kinds of mistakes learners make during recall fall neatly into three kinds: (i) the learner incorrectly applies a learned signal to a different meaning and produces a novel mapping between meaning and signal (this is why the language becomes underspecified in both conditions); (ii) the learner confuses parts of different signals and combines them to produce a novel signal (this frequently happens early on in both conditions); or (iii) the learner makes a typographical error and produces a novel signal (this can theoretically happen at any point, but does not typically occur at all if the number of distinct strings is very low). The latter two error types are, by their very nature, low frequency additions to the system. This means that if they occur in the semantic bottleneck condition, they are extremely likely to be removed from the system almost as soon as they are created.

Of course, this is only half of the story however. We also only see participants making mistakes that result in the innovation of new signal forms when there are enough signal types in the system to cause difficulties in learning. In the nSB condition, signals are lost so quickly early on that it soon becomes unlikely to even make a mistake in the first place. This can be seen by examining the point at which new signals stop being created in the two coalescent trees (i.e. where we see the last relationship of possible descent). In Figure 4.6 we find new signals being created even in the final few generations in the nSB condition, whereas in the SB condition (Fig. 4.4) this ceases by the third generation. The same seems to hold true of the other chains in both conditions: if we note at which point new signals stop being generated we find that this occurs between generations 5 and 10 in the nSB condition, whereas for the SB condition it occurs earlier, between generations 3 and 7.

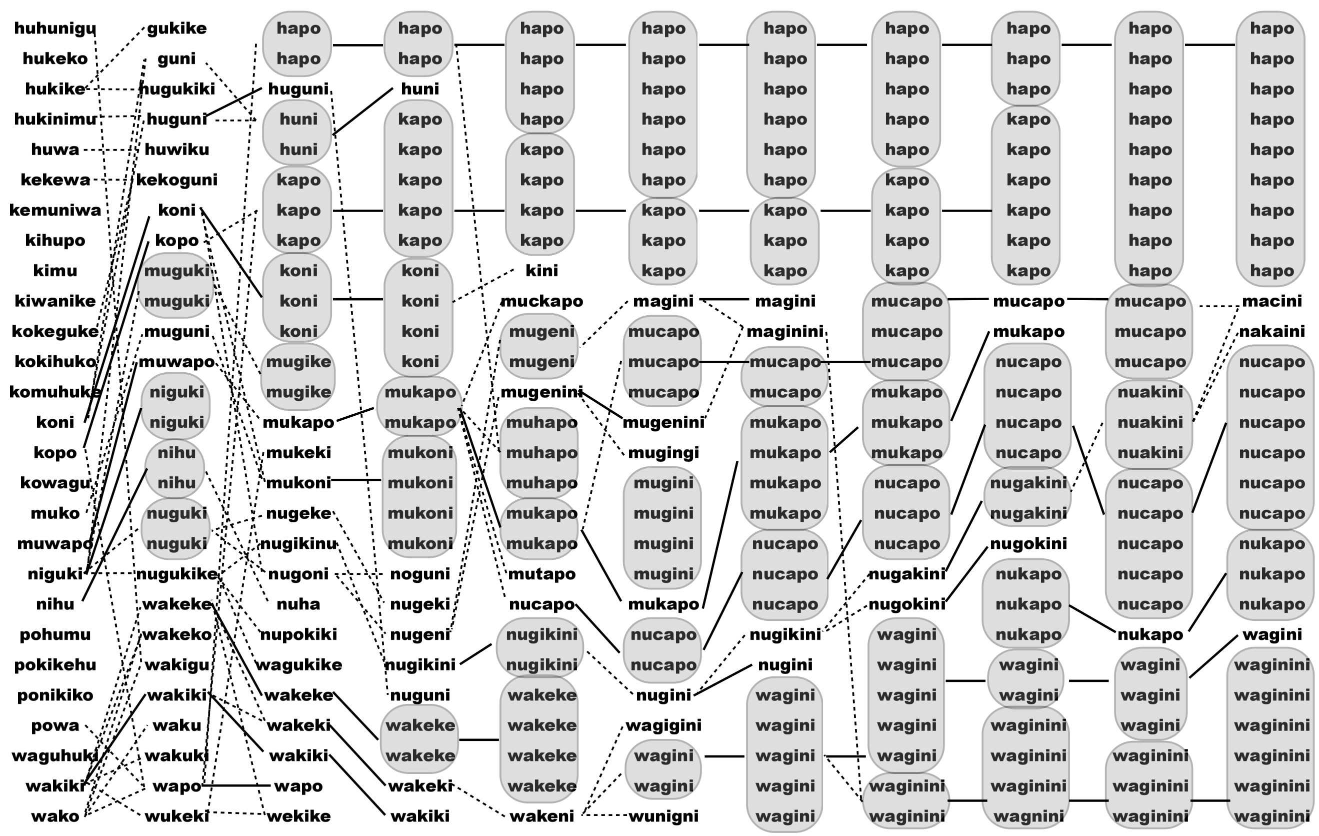


Figure 4.6: A coalescent tree showing relationships (solid lines) and possible relationships (dotted lines) of descent between signals in chain A of Experiment II. In this condition, all signals were seen by the next generation. Repeated signals are grouped together for clarity. This tree demonstrates how signals evolve over time. Although the amount of variation in signal types does decrease over time, the formation of new signals is seen even in the final few generations.

4.3.3 Summary

We can summarise these results as follows. Firstly, just as in the previous study, we can confidently accept both the learnability and the structure-increase hypotheses on the basis of significant decreases in transmission error between languages, and significant increases in the amount of structure found in each language over time. Furthermore, we can also accept the systematic underspecification hypothesis, as languages in this experiment adapt to be learnable by encoding fewer dimensions of meaning. However, in spite of these similarities, there were also some differences between the languages constrained only by limitations in human memory, and those constrained by an external semantic bottleneck. First among these is the stronger persistence of variation -- both in terms of a greater number of distinct signal types, and also irregularity. Secondly, one of the chains evolved a language that was not only systematic in the way it was underspecified, but also had internal structure that was regularly aligned with some aspects of meaning. Suffixes regularly expressed different types of motion (spiralling vs. movement-to-the-right), and for some meanings, shape was locally encoded as a prefix.

4.4 Discussion of Experiments I and II

This section discusses the implications of the results of the first two experiments in a little more detail. The key point to note is that we see structure emerging as an adaptive response to any kind of bottleneck on the process of transmission. These studies extend previous computational results by demonstrating that the presence of a semantic bottleneck is not essential in order for cultural transmission to become adaptive (Cornish, 2010). Instead, all that is required for adaptation is imperfect information. Whilst the actual source of that imperfect information (i.e. natural limitations on human memory, or restricted exposure to data encountered in the world) is unimportant for the emergence of structure *in general*, it might play a role in shaping what that structure looks like. In other words, the dynamics of cultural transmission are as important as the mechanisms constraining it in accounting for the emergence of a specific system.

As a case in point, we have seen how the loss of variation early on inhibits the generation of signal innovations. Whilst this does encourage more stability in the resultant languages, it also leads to features which seem less useful for the task of communication (i.e stability in this case comes at the expense of being able to unambiguously identify more of the meanings in the meaning-space). In short then, one of the main findings to emerge from these two studies is that the principle of linguistic adaptation holds true in a population of human learners. Structure is the inevitable result of transmission with constraints. Whatever emerges will be adapted to those constraints, but perhaps more tellingly, to those constraints *only*. Given that there was no explicit pressure on the systems to create unambiguous mappings between meanings and signals, we should perhaps not be too surprised to find that systems with those qualities do not naturally 'fall out' of the process. This is an issue we will be turning our attention to in the next chapter.

The second point to note concerns the differences between the two experiments: although the languages in both studies are adapting to fit the learning biases of the human participants, the semantic bottleneck in the first experiment additionally inhibits the survival of signals at random. This has some interesting implications. In the second experiment, the only processes occurring during transmission are selective -- for greater salience, learnability, cohesion with other words, relationship to the meanings, etc. Whilst this kind of cultural selection of signals fitting these constraints is still present in the first experiment, the effect of the semantic bottleneck is subtly different.

In some sense, the action of the semantic bottleneck equates to what biologists would call neutral selection without innovation. This process has been used in biology to explain how traits in a population can rapidly change over time and fixate on one particular variant, without that variant being connected with any enhancement of fitness for the organism (Nettle, 1999). Rather than being about Herbert Spencer's term, 'survival of the fittest', neutral drift is about 'survival of the luckiest' (Kimura, 1989). Models have shown that given a population of variants that are not under selection for any particular trait, and a random policy of removal of those variants steadily over time, taking this process to its logical extreme will result in one variant taking over entirely (Cavalli-Sforza, 2000; Nettle, 2009). In fact, this is very nearly what we found in one of our chains (A) in the first experiment.

In sum, whilst there is there is a only a one stage process of cultural selection (for, amongst other things, signals that are more easily acquirable) in Experiment II, there is a two stage process in Experiment I. First the semantic bottleneck removes items at random, and then cultural selection prunes what is left. This explains the difference in the amount of variation in each condition.

How are we to interpret the findings here? In some sense it is disappointing that we did not find more evidence of compositional structure here. That is not to say, however, that the structure we did find is not in itself 'language-like'. In fact, we find that underspecification is actually rife in natural language. As an example, we can consider the case of common nouns. Unlike proper nouns which pick out unique referent in the world, common nouns refer to entities of a general type. Something similar appears to be happening in the languages arising over the course of the transmission chains in these experiments. Rather than picking out every potential meaning to be distinguished, it seems that more general categories of meanings are forming. In a sense, the signals are acting as a cue to participants to lump together meanings with similar features[15]. This (re)categorisation of the meaning-space appears to be part of what gets transmitted to the next generation.

---

[15] Recall the comment from one of the learners in the first experiment about how the aliens did not seem to see colour (§4.2.2).

In order to explore how language and cultural transmission interacts with categorisation in more detail, Matthews (2009) (see also Matthews, Kirby & Cornish, 2010) used this experimental framework to investigate how signals evolve in a world with continuous meanings. Instead of having discrete, finite meanings, this study explored what would happen if human participants were trained on signals paired with meanings whose features varied along continuous dimensions. Using morphing software, the dimensions of shape and orientation were manipulated to create a meaning-space where horizontal triangles mutated into vertical rectangles, and vertical triangles mutated into horizontal rectangles. This created 100 different meanings, each of which was only very slightly different to its neighbouring meanings.

Initial languages were randomly constructed, and participants were given a subset of training items to learn, drawn from the meaning-space at random. During the test phase, they were asked to provide signals for another subset of meanings, again drawn at random. These meanings were almost certainly not the same as those that they had seen in training, yet nevertheless, their output was recorded and transmitted to the next generation. Using a control set of items that participants were always asked to name after they had produced the new language for the next learner, Matthews (2009) demonstrated that the signals began to partition the meaning space up in different ways. The languages in this study were structured by categorising similar meanings with similar strings. Furthermore, there was interesting variation between transmission chains as to how this partitioning was achieved. Not only were signals and category boundaries transmitted between generations, but the very notion of how 'similarity' was defined also got culturally transmitted. For instance, some chains were blind to the rotation of the objects, and would classify objects as similar if they were the same basic shape in a different orientation, whereas other chains classified objects in different orientations as being dissimilar.

It is interesting that in Experiments I and II we find that structure also emerged via categorisation. As categorisation underlies linguistic properties like compositionality and recursion, this makes these initial results appear a little more encouraging. Taking another perspective, let us think for a moment about how well the systems that emerged would function in a communicative context. Participants at the beginning have very little chance of successfully communicating any of the meanings to another learner. Participants at the end would be able to convey quite a few meanings reliably, but certainly not all 27. Although we did not find robust signs of compositionality emerging like the simulations did, these results are at least suggestive of the idea that features of language that are useful for communication can emerge from transmission constraints alone.

Chapter Five

Language Adapts to be Expressive

The first two experiments described in the previous chapter have shown that language adapts to all and only those constraints being placed upon them[1]. In these studies, the only pressure being put upon the systems is to be learnable. As a result, the languages adapt in ways that suit this outcome. However, some of the structural features that emerge from this process, such as the widespread underspecification of the meanings, although well attested in natural language, do not seem particularly suited for communication. A potential reason for this stems from the fact that there is no pressure constraining the languages to uniquely express all of the meanings. This raises an interesting question: if we build in a pressure for expressivity, will we find compositional structures better suited for communication emerging? This is what the experiments in this chapter aim to find out[2]. Before we begin, we should perhaps spend a moment thinking about what we mean by 'expressivity'.

Expressivity relates to the ability of signals to differentiate meanings within a language. As such, it is also related to the amount of variation in a system, and also

---

[1] This includes not only the external constraints being manipulated in the experiment itself (i.e the semantic bottleneck), but also those cognitive constraints *internal* to the learner.

[2] The results reported in this chapter have also appeared in several publications, including Cornish (2010), Cornish *et al.* (2009), Kirby *et al.* (2008a) and Kirby *et al.* (2008b).

to the number of distinct signals. It is different to the kinds of distinctiveness requirements hypothesised to drive song creativity in certain species of birds (Marler, 1957; Ptacek, 2000), because that implies that the signals themselves must be dissimilar from one another or from the signals produced by other vocalisers. This is not the case with an expressivity requirement: here, all signals must do is be able to express all of the meanings that a speaker wishes to convey in such a way that a hearer can easily recover those meanings. We might, therefore, expect a language which is fully expressive to have one-to-one mappings between signals and meanings, and contain little or no redundancy or ambiguity[3].

In the case of natural language, the ability for signals to unambiguously differentiate between possible meanings comes directly from the pressures of communication itself. Since one of our aims is to explicitly investigate whether we can observe the nonintentional emergence of language-like structures in humans, we will have to find some other way of encouraging expressivity in our experiments. The next section discusses ways we can possibly achieve this.

5.1 The Expressivity Requirement

If we take a closer look at various iterated learning models, it turns out that they all have an expressivity requirement built in somewhere[4]. Even in ILMs which purport to have no meaning-space at all (e.g. Teal & Taylor, 2000; Zuidema, 2003), the algorithms insist that agents must continue creating signals until they have satisfied

[3] In reality, we know that human language is not nearly so perfect. Given that communication occurs over a noisy channel and involves making generalisation inferences over imperfect data, some degree of ambiguity and redundancy is actually hypothesised to be adaptive, and perhaps even the inevitable outcome of processes of iterated learning (Hoefler, 2006). Also, we should not underestimate the role of pragmatics and context as a disambiguator in real world communication.

[4] See Cornish (2005) for more discussion of this phenomenon, and an example of a small ILM study that shows how even initially highly structured and varied input degrades rapidly without the pressure to maintain a minimum degree of expressivity.

a minimum level of expressivity. If we take the model by Zuidema (2003:55) for instance, we find that this requirement is controlled by a parameter, *E*:

> "To avoid insufficient expressiveness, we also extend the generalization step with a check if the number E_G of different strings the grammar *G* can recognize is larger than or equal to *E*. If not, E-E_G random new strings are generated and incorporated in the grammar."

Examining the results, Zuidema (2003:56) further notes that "after an initial phase of over-generalisation, the expressiveness remains close to its minimally required level". Let us contrast this for a moment with what was going on in Experiment I. Humans, unlike the computer agents, were given only one chance to produce a signal for a given meaning. They were not forced to continue to produce an output until they had produced 27 distinct signals.

In some ways it is surprising that participants do not seem to do this naturally. Several studies have investigated the claim that humans have a one-to-one mapping bias, and found it to be rather robust (e.g Slobin, 1977; Haiman, 1980; Macnamara, 1982). When children learn a new word, studies have revealed that they make several assumptions when trying to identify the correct referent from a context. In addition to preferences for whole, rather than parts of objects, attention goes initially towards those objects which are unnamed: children never instinctively assume that a referent has more than one name (Markman & Wachtel, 1988).

In a discussion on learning biases implemented within the ILM, Smith (2003) makes explicit mention of one-to-one mapping biases: the models do not naively include an expressivity constraint by accident, it is there to model known human biases. The results of the experiments in Chapter 4 are therefore potentially of wider interest as they reveal that participants are happy to over-ride this learning bias given the right circumstances during cultural transmission.

5.2 Experiment III: Adding an Expressivity Constraint

There are many ways we could go about introducing an expressivity constraint. One method that we have already discussed would be to force participants to keep producing utterances until they have produced 27 distinct ones (§5.1). There are several reasons why this is not an ideal way to enforce expressivity. The first is methodological. How do we decide which meanings to show to participants again? Do we show them all? What if a participant responds to a given meaning differently each time we present it? How do we decide which of the signals to transmit to the next generation in this case?

The second issue is one of preserving the integrity of participants belief in what the task is about. We want to make sure that participants are unaware that they are changing the language[5]. We would have to find some way of justifying why we were continuing to prompt learners for responses if we wanted to maintain this illusion. Of course, we could just tell participants that they had used the incorrect signal, but this leads to more methodological issues. If a participant makes a genuine mistake and uses the wrong signal for a meaning early in the test, when the real meaning appears later on and they try to use the signal that they know to be correct they will be (falsely) told that they are wrong. This would be both confusing and demoralising. Note we cannot just keep prompting them until they produce exactly the same responses as the previous generation, as (a) this would lead to no change in the system, and (b) they may only have seen half the data.

The effect that we are finding when there is no pressure for expressivity is that it leads to an increase in the number of homonyms in the language: signals become

---

[5] This goes back to our attempts to design an empirical framework suitable to investigating what Croft (2000) describes as nonintentional mechanisms of change (§3.3). If participants believe the task is about anything other than reproducing the data they have been given as faithfully as possible, then we can no longer be sure we are seeing changes that : "happen as a consequence of the act of production" (Croft: 2000:65).

associated with more than one meaning, and this leads to ambiguity in the system. Although it turns out that this is exactly what we should predict given the transmission constraints imposed, if we want to model the emergence of an expressive language we need to find a way to handle the ambiguity that this introduces. We have actually already seen another way to handle the introduction of ambiguity in computational simulations of iterated language learning (§4.1.2). In Kirby (2001), ambiguity is prevented from accumulating in the system by the learning agents being programmed to ignore any signals for which they already have meanings for. If agents later see the same signal paired with another meaning, they are prevented from adding that mapping to their grammars. Obviously, we cannot prevent our human learners from associating signals with multiple meanings in this way. We can, however, ignore repeated signals on our participants behalf by choosing to not select them during transmission (see §3.4.2). In other words, we cannot prevent learners from introducing ambiguity into their output, but we *can* prevent it from featuring in their input.

We can refer to this process as **filtering** (see Kirby, Cornish & Smith, 2008). What this amounts to is the idea that learners only learn from novel signals. Note that this is actually quite a realistic assumption: if a learner hears a signal that they think they already know the meaning for, they are unlikely to actually check to see whether the signal actually matches up with the meaning that they believe is intended. They will parse it and move on. If they hear a novel signal on the other hand, they will always check to see what it could possibly relate to in the world. Due to the way meanings and signals are always presented together, participants in the studies are always made aware of any underspecification of the meanings, therefore intervention to correct this is, I believe, justified[6].

[6] It would be interesting in the future to test whether this is the case using this experimental framework. One manipulation we could explore might involve giving participants the choice of whether to 'inspect' a meaning or not. If they were also given either a time-limit or a limit on the number of inspections they could make over the course of learning, we would predict that learners would limit their choice of when to check up on the meaning of a signal to only those cases where they encounter a novel utterance.

The way the filtering process was implemented was as follows. experiment proceeds as normal until the transmission phase. Recall that it is at this point that meaning-signal pairs are selected to be transmitted to the next generation. In the case where there is no semantic bottleneck in place, this phase involves no real selection or transformation of the data[7] -- the entire output simply gets passed on from the previous learner. When there is a semantic bottleneck in place however, a subset of meaning-signal pairs are randomly chosen to be transmitted. It is after this selection procedure that the expressivity filter applies. If any training items contain homonymous signals, all but one (selected at random) is removed from the subset to be given to the next learner.

The expressivity filter does create one issue; namely that the removal of homonyms after the semantic bottleneck has been applied means that participants are no longer guaranteed to receive 14 items in their training input. Two strategies were considered for ensuring that the number of training items remained constant, but ultimately, both were rejected. The problem is that they introduce selection biases into the experiment. For example, the first strategy that was considered involved removing homonyms from the language *before* the bottleneck applied. This plan was rejected for two reasons. Firstly, it would not work when learners had fewer than 14 unique signals in their whole output. Of course, if this was the case, one suggestion might be to use this criterion (at least 14 unique signals) as a benchmark for rejecting that participant, and try training a new learner on the same data in hopes that they might recall more items. Note that this is equivalent to selecting learners with above average recall ability, which is something that we did not do in our previous experiments.

The second reason for rejecting the strategy of applying filtering before the bottleneck relates to the effect it has on the selection of meanings. Currently

[7] Technically, even when there is no semantic bottleneck in place there is still selection of training-items, which ensures that the new learner receives them in a different order to that in which they were produced.

meanings are selected at random. This models the idea that we are motivated to communicate about things that just happen to be going on in the world around us. If we switch the order of the two processes (filtering and selection of meanings to convey) then the world is no longer our random guide. The world is suddenly under pressure to only present us with a subset of meanings that are related by the (fairly arbitrary) fact of having a unique signal. To clarify, when filtering follows the bottleneck, every meaning is equally likely to be selected to be transmitted. if filtering precedes the bottleneck however, some meanings are more likely to be selected than others simply because they are expressed with a novel signal. As appearing in a learner's training input is highly correlated with being faithfully acquired (and therefore, with the preservation of the distinctiveness of the signals), these meanings are likely to be continually sampled again, and again, and again. Eventually we will just see the transmission of the same 14 meanings with just minor variations in their signals.

The second strategy that was considered does not fare much better. This idea involved actively replacing repeated signals with a novel one. Two options for implementing this were debated. The first was inspired by ILMs like Zuidema (2003), who had agents randomly construct novel signals when they had run out of unique learned signals. The first issue with this approach is that it is clearly not the natural response of our learners. It is therefore unclear what creating a random signal string corresponds to in real-life. The second issue relates to the fact that the signals in a given system at any point in time are there because they share a common history. They have evolved together and have the features that they do in virtue of this fact[8]. If we start introducing completely novel signals at later points, those signals will, (i) be recognisably different, and (ii) have a potentially disruptive

[8] We have already seen this clearly in the coalescent trees in the previous chapter (§4.2.2 and §4.3.2), which show how signals evolve over time as a result of descent with modification.

influence on the language[9]. It is as if we were undoing all of the evolution of that particular meaning-signal pair, and resetting it to generation zero. A less extreme alternative that was considered was to replace the signal that was filtered out with the last novel signal that a previous participant in the chain had produced[10]. In the worst case scenario this is equivalent to generating a new random signal as it might involve going back to the original form in the initial language. However, more often than not this might involve only stepping back a few generations. Whilst this would be less noticeably disruptive than random invention, it still has the problem of destroying lines of inheritance.

In summary, the two options that I considered for ensuring that participants always received 14 items in their training input were to either (a) apply the filtering process before the bottleneck, or (b) to replace those repeated signals that got removed with an alternative. Both of these strategies were found to introduce selection pressures not found in the previous experiments; for either learners with better recall, or for just a subset of meanings having distinct signals; or else they interrupted the evolution of signal forms themselves. Given this, it seems that the correct course of action is to accept that learners might sample from fewer meaning-signal pairs during training. Not only does this more closely replicate processes occurring in some ILMs (e.g. Kirby, 2001), but it also maintains greater continuity with the previous experiments. Both this and the fact that the filtering technique also has the added advantage of being invisible from the point of view of the learner, will be especially important for comparing this study to the studies in the previous chapter.

5.2.1 Method

---

[9] The results of Zuidema (2003) show that given an extremely large number of generations, these issues will 'wash out' over time. However, given that we have practical limitations on the number of generations we can run, these concerns are important.

[10] To clarify, we mean that the signal is novel with respect to the current generation, and not necessarily novel with respect to the generation of origin.

The same basic methodology described in §3.4 was used in this experiment, however, this time we added in a pressure to express meanings uniquely. Because the experiment with the semantic bottleneck was slightly more economical to run in terms of time, and most ILMs include a semantic bottleneck, we use one here as well.

Aims and Experimental Hypotheses

We want to find out whether adding an expressivity constraint to the task encourages compositional structures to emerge. Our hypotheses are therefore almost identical to those in the first experiment (§4.2), but with one slight modification: instead of predicting that compositional structure will emerge as the result of pressure to generalise to novel stimuli, it is predicted that we also need a pressure for expressivity. The three hypotheses have been reproduced below.

1. **The Learnability Hypothesis**: Languages will become easier to learn as a result of iterated learning.
2. **The Structure-Increase Hypothesis:** Languages will become more structured as a result of iterated learning.
3. **The Compositionality Hypothesis:** Pressure to generalise to novel stimuli *combined with a pressure to uniquely express each meaning* will result in languages evolving to become compositional.

Experimental Design

Four initial languages were randomly generated to seed each chain[11]. Each chain was run for ten generations, and had both a semantic bottleneck in place as well as filtering. This meant that participants were trained on around 50% of meaning-signal pairs and all meaning-signal pairs represented a unique one-to-one

[11] One of the chains in this experiment (D) was actually obtained during my MSc project, where this framework was piloted (Cornish, 2006).

mapping[12]. The training and test procedures were identical to those outlined in Experiment I, with three rounds of training, each followed by a test round. Only the signals produced for meanings in the final test were transmitted to the next learner. The experiment was run using E-Prime software, and the results were analysed using *R*.

Participants

Another 40 participants were recruited via an advertisement at the University of Edinburgh's student employment services to take part in this study, of which 18 were male and 22 were female. They had an average age of 22.75 (*S.D* = 4.53), and were each offered £5 for their participation. Participants were assigned to a generation and chain at random, and had normal or corrected-to-normal vision, were fluent English speakers, not dyslexic, and had not participated in any of the previous studies or taken courses in linguistics before. The study conformed to the ethics guidelines set by the University of Edinburgh's College of Humanities and Social Science, and participants were fully briefed before taking part of their rights to withdraw from the experiment at any time, and that their participation and results would remain anonymous.

Procedure

The training procedure was identical to that of the first experiment. Training items were displayed in a random sequence on a monitor for six seconds each. Every training item was seen six times over three rounds of training - twice per round. During testing phases participants were shown meanings and given an unlimited time to provide the correct signal associated with that meaning. No explicit feedback was given to participants during training or testing, and participants remained unaware of the true purpose of the study until after they had completed

[12] A retrospective analysis revealed that on average, participants in this study were actually trained on 12 items (44% of the data), instead of 14.

training and testing, filled out a questionnaire about their experiences, and been debriefed about the true purpose of the experiment.

5.2.2 *Results of Experiment III*

Structure and Learnability Increase

We can apply our measures of structure and transmission error once more to get a general overview of what is happening to the languages over time. As Figure 5.1.left demonstrates, we find that all four chains show significant levels of structure by the second generation, and a general trend for these levels to increase over time thereafter, whilst Figure 5.1.right indicates a steady decrease in the amount of transmission error between learners over time. Paired t-tests on the structure scores obtained in the initial and final generations confirm that the increase was statistically significant (as shown by a mean increase in structure of 6.77, t(3) = 2.535, P<0.04), and examination of the transmission error scores reveals a similar story. Languages are significantly more learnable by the final generation than they are at the start (mean decrease in error of 0.43, t(3) = 8.056, P<0.002).

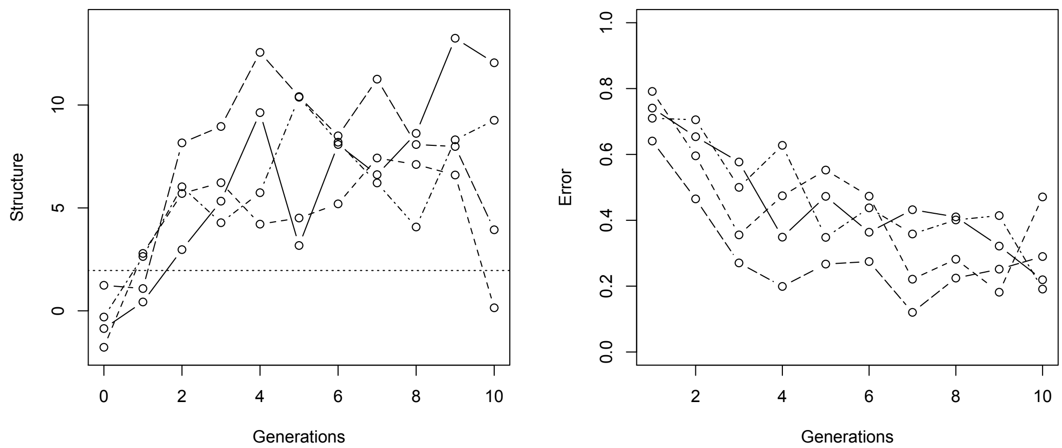


Figure 5.1: Structure and learnability increase over generations in all four transmission chains in the filtered condition. Languages become significantly more structured over time (left) , whilst

transmission error between adjacent generations significantly decreases (right). Redrawn from Kirby, Cornish & Smith (2008) using 10,000 randomisations instead of 1,000.

The language produced by the final generation of chain B is notably an exception to this general pattern. Not only are the levels of structure negligible for this learner, but examination of the error scores (Fig. 5.1.right) reveal a sharp spike, indicating that the language has undergone a lot of change. Visual inspection of the language in question suggests that a possible reason for this is a combination of the random application of the semantic bottleneck, which removed a few of the more key exemplars required for inferring the structure of horizontally moving objects in particular (see Appendix B3), and a slightly poorer than average recall of seen items by the individual learner in question. As an example, the learner at generation 10 reproduced only two items perfectly, compared to the previous three learners who on average reproduced six items perfectly. Individual variation in recall ability is to be expected of course, although we should predict that as chains are run for longer, this should have less of a de-stabilising impact.

Pressure for Expressivity Increases Signal Types

As mentioned in the introduction to this section, our pressure for greater expressivity is indirect and quite subtle. One thing we need to determine is therefore whether our manipulation has actually had any noticeable effects. In other words, does our filtering process actually encourage the preservation of more distinct signals? In order to judge this, we can examine the number of distinct signal types, and compare them to those obtained in the unfiltered condition (Exp I). Figure 5.2 shows this information in a box-plot. For ease of comparison, the data from the unfiltered condition has been reproduced in the same format[13].

From this we can see that although there is some initial loss in the filtered condition (Fig. 5.2.lower), the overall level remains fairly stable across all four chains. This

[13] This information was previously shown in Table 4.2.

stands in strong contrast to the unfiltered condition (Fig. 5.2.upper), where the loss is both more severe, and increases cumulatively over time. In order to determine whether the differences between the number of signals in both conditions was significant, an unpaired one-directional t-test was performed at each generation. The levels of significance have also been indicated in Fig 5.2 (*n.s* = not significant, * = P<0.05, ** = P <0.005). These reveal that from the second generation onwards, the number of distinct signals in the filtered condition was indeed greater than that obtained in the same generation in the unfiltered condition. Therefore we can conclude that filtering is having the desired effect of encouraging more of the meaning space to be uniquely expressed with a single signal.

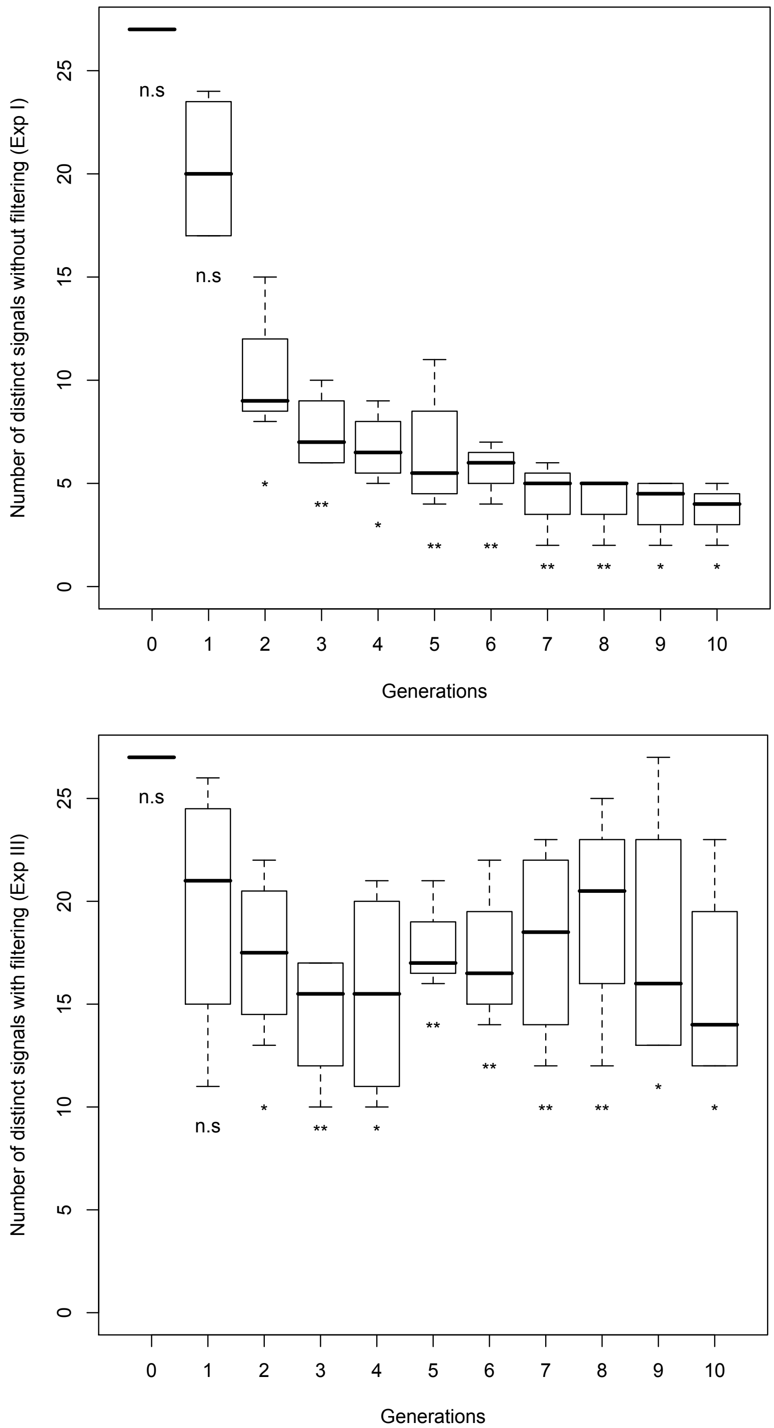


Figure 5.2: Box-plots contrasting the number of distinct signals found over generations across all four chains in the filtered (Exp III) and unfiltered (Exp I) conditions. Horizontal lines indicate the median number of signals, boxes indicate the interquartile range, and whiskers indicate maximum and minimum values found. Whereas the number of signal types in the unfiltered condition (upper) rapidly decreases over time, those in the filtered condition (lower) remain relatively stable and high.

Significant differences between the two conditions were found from generation 2 onwards. Taken together this shows that filtering is indeed encouraging greater expressivity in the languages.

Compositionality: an Adaptation

Looking at Fig 5.2.lower more closely, we find that at generation 9, at least one of the languages actually has 27 distinct signals. If we examine this language in more detail, we find that it is in fact, compositional. Table 5.1 shows what this system looks like. Each cell in the table contains the signal associated with each of the 27 meanings, with columns corresponding to a colour (black, blue or red), whereas horizontal rows reflect motion and shape according to the symbols shown. We can see that signals in this language are composed of three segments, each of which conveys a different aspect of the meanings. The colour of the objects is regularly encoded in the first segment (which corresponds to the initial letter: **n-** for black; **l-** for blue; and **r-** for red) whereas motion is regularly encoded in the final segment (**-ki** for horizontal; **-plo** for bouncing; and **-pilu** for spiralling). Shape is more unreliable, but is otherwise encoded in the middle segment (there is a general tendency for squares to contain **-re-**, **-ne-** or ∅; circles to contain **-ho-** or **-he-**; and triangles to contain **-ki-** or **-ke-**).

The one exception to this pattern is the signal for a horizontal red square - **renana** - which appears to be an irregular. Examining the history of this chain we find that by chance, this meaning-signal pair was seen by every generation from generation 8 onwards. This ties in nicely with the findings of Kirby (2001), which showed that irregularity could arise and be maintained in systems transmitted via iterated learning provided that some meanings appeared more frequently than others. Although this meaning certainly did not appear in participants training any more than any other, the fact that it reliably survived the bottleneck at a time when therest of the system was undergoing changes to become more regular and rule-based has allowed for it to retain its idiosyncratic structure. This is of course, just a single anecdotal case and further investigation using this experimental framework, along the lines of the study conducted by Beqa *et al.* (2008), is required[14].

nereki	lereki	renana
neheki	lehoki	reneki
nekeki	lakeki	raheki
nereplo	laneplo	replo
nehoplo	lahoplo	rehoplo
nekiplo	lakiplo	rahoplo
nepilu	lanepilu	repilu
nehopilu	lahopilu	rehopilu
nekipilu	lakipilu	rahopilu

Table 5.1: A fully compositional language arising from Experiment III (Chain A, generation 9). This language has 27 distinct signals for each of the meanings, and each signal is composed of three segments. The first segment represents the colour, the middle segment represents the shape, and the final segment represents the motion of the object. Reproduced from Kirby, Cornish & Smith (2008) with permission.

A similar compositional system to this also emerged in Chain D in this condition, between generations 4 and 7, although it was not nearly as perfect as the example shown in Table 5.1 (see Appendix B3). This is encouraging as it indicates that this is

[14] This study follows the same methodology outlined here but specifically manipulates the regularity of initial variants (in the language given to the first learner, half the signals were regular, half were irregular), and the frequencies that certain meanings appear in during training (creating a set of high-frequency items, and a set of low-frequency items). It demonstrates that low-frequency items are much more likely to become regularised over time if they were originally irregular, than high-frequency items were. In effect, this study successfully replicates the Kirby (2001) findings but uses human learners instead of artificial agents.

not just a peculiarity of this chain alone. The fact that compositional systems, when they do arise, do not seem to be very stable is intriguing however. We will come back to this in Experiment IV.

The Evolution of Signal Segments in a Compositional Language

The fact that we can decompose the final few generations of this language into different signals segments (i.e. beginning, middle and end) is interesting. How might this segmentation structure have arisen historically over time? In order to address this question we can examine the relationships of descent between signals in a coalescent tree, like we did in the previous two experiments. Unlike in the previous two experiments where we analysed the signals as holistic units, and examined how forms changed and increased in frequency over time, this time we can analyse the signals by segments (Cornish, Tamariz & Kirby, 2009).

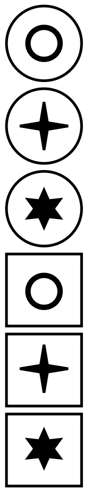
We are primarily interested in quantitatively determining the extent to which the signals have adapted to the structure of the meaning space, and when that might have occurred. Cornish *et al.* (2009) implement a technique for doing this using this data[15]. The first step in this process relies upon us being able to make a parsimonious segmentation of the signal-strings in the language into elements that correspond to different aspects of meaning. We need to be able to examine the language in its final generation and formulate rules like "the beginnings of signals consistently encode colour" or "signal-final '-pilu' reliably encodes the motion of spiralling". If we do this to the language shown in Table 5.1, we end up with each string being divided up into three different sub-strings as described earlier. In order to allow for a consistent analysis, we then need to carry this segmentation pattern back to all previous generations.

---

[15] The Cornish, Tamariz & Kirby (2009) paper represents joint work, and will be referred to several times in this chapter. To be clear: H.C. and S.K. designed the study. H.C. collected and analysed the language data. M.T designed and performed the *RegMap* and coalescent tree analysis. H.C., M.T., and S.K. contributed to writing the paper in that order.

Before we begin, we need to introduce some terminology to enable us to describe what is going on. An analysis of the language consists of determining the following: (i) **signal segments** - the position within the string where different meanings are conveyed (in this case, the beginning, middle, or end of words); (ii) **signal segment variants** - actual tokens residing in a given segment position (e.g. '-pilu', 'n-', or '-aho-'); (iii) **meaning elements** - aspects of meanings (i.e. the features of the meaning space, like motion, shape and colour); and (iv) **meaning element variants** - actual instances of a given meaning (i.e. the values of particular features of the meaning space, like 'black', 'bouncing', or 'triangle') (Cornish *et al.,* 2009).

Fig 5.3 illustrates how such a segmentation process might occur using a toy example with signals associated with a meaning-space varying in two features (shape: circle or square) and three values (shape of insert: circle, cross, or star). Beginning with the most recent generation (4 in this case), the signals are analysed into two signal segments: the first indicating shape of object, the second indicating the shape of insert. We then look to the immediately preceding generation and do the same thing, keeping in mind both what the previous signal segment variants were (i.e. DO, RE, MI, FA, SO) and that we must always have two segments. We aim to always preserve signal variants from the more recent generation, unless a more parsimonious analysis presents itself.



Meaning	Generations 4	3	2	1
	DO.MI	DO.MIR	DO.MIR	DO.MIR
	DO.SO	DO.LA	TI.LA	TI.X
	DO.FA	SO.FA	RE.FA	DO.FA
	RE.MI	RE.MI	RE.MIR	RE.MIR
	RE.SO	RE.SOR	RE.SOR	RE.SOR
	RE.FA	DO.FA	DO.FA	X.FA

Figure 5.3: An example of the segmentation process at work. Strings in the final generation (4) are segmented into two parts according to the most parsimonious alignment between meanings. This segmentation pattern is then carried back to earlier generations one at a time. Criteria for determining where segmentation boundaries are must take into account both (i) the segmentation patterns seen in the more recent generation, and (ii) the most plausible segmentation within the current generation being processed. Segment locations are represented by a full stop, and null segments by X.

Applying this segmentation procedure to generation 2 results in the identification of two new signal variants (MIR and SOR). At generation 3 we are presented with the novel string 'TILA'. Due to the fact that we treated 'LA' as a suffix in the previous generation, and all other prefixes in the signal space appear to be composed of two letters, we segment the string as 'TI.LA'. It should be clear that at some point we will get to a situation where we have to posit a 'null' sub-string in place of one of our signal segments (in the example above this is represented by X). In the toy example, this happens twice at generation 1, where we find the signals 'TI' and 'FA'. Do we posit a null signal variant in segment position one or two? In this case, we must view each case on its own merits. Given both the segmentation suggested by the previous generation as well as the presence of 'DO.FA' in the current generation, it makes sense to posit the null variant before 'FA'. In the case of 'TI', we have no evidence supporting it as being a prefix or a suffix in the current generation.

However, in the previous generation it was analysed as a suffix, so it is more parsimonious to do the same again here.

Cornish *et al.* (2009) applied this segmentation technique to the compositional language described earlier. Using this, they examined how the lineages of signal segment variants appearing in the final segment position changed over time[16]. Figure 5.4 is a reproduction of this coalescent tree[17]. From it, we can see a similar pattern of emergence to what we saw previously when examining whole signals. Early generations contain many low-frequency signal variants, which quickly reduce in number until just a few high-frequency signal variants remain. These variants which appear at each generation are not random, but are related to those that appear before. In some cases this involves direct and perfect replication of a variant (indicated by solid lines), but even where new signal variants appear, it is easy to determine possible relationships of descent between seen variants (indicated by dotted lines).

Often, many of the changes that occur to form new signal variants appear similar to those that are well attested in natural language change. For instance, Cornish *et al.* (2009) note that we find cases of phonological reduction ('hona' becomes 'na'), metathesis ('neki' becomes 'nike'), single segment replacements ('pilu' becomes 'pilo'; 'nepi' becomes 'napi') and blends ('humo' and 'huna' combine to form 'homa' and 'hona'; 'na' and 'ki' merge to form 'neki'). They also point out that the frequency of each signal variant, over time, comes to confirm what we already know: that the language is adapting to a meaning-space consisting of three elements. This can be seen most readily at generation 4, where we find just three variants ('na','neki' and 'pilu') each appearing nine times , and again in generations

---

[16] Note that we can also do the same for variants appearing in the initial and middle positions.

[17] This coalescent tree differs from those shown in Chapter 4 in that frequency information is indicated in brackets. This is a purely cosmetic difference, making it easier to spot frequency patterns between generations.

9 and 10, where we find ‘plo’, ‘pilu’ and ‘ki’ (with the irregular ‘na’, being part of the renana example discussed earlier).

From this, we can hypothesise that signal endings came to perfectly encode one of the meaning elements in generation 4 (we do not know which at this point, as we have not been factoring meanings into our analysis), and again in the final two generations. This is interesting, as if we return to Fig 5.1.left, our measure of structure shows this chain at this generation to have only moderately high levels of structure. There is no sign in the Mantel test to indicate that we have a perfect structural mapping between parts of the signal and parts of the meanings. Of course, if we wish to see whether this occurs at other points in the chain we can also draw up coalescent trees for the remaining signal segments (i.e. the beginning and middle positions) in order to determine when other forms arose to encode meaning elements. However, it would be better if we could determine this statistically. Fortunately, the fact that an analysis of the frequencies of signal variants alone provides clues as to how meanings break-down into elements, is something that we can exploit if we wish to analyse the emergence of compositionality more quantitatively (Cornish *et al.*, 2009).

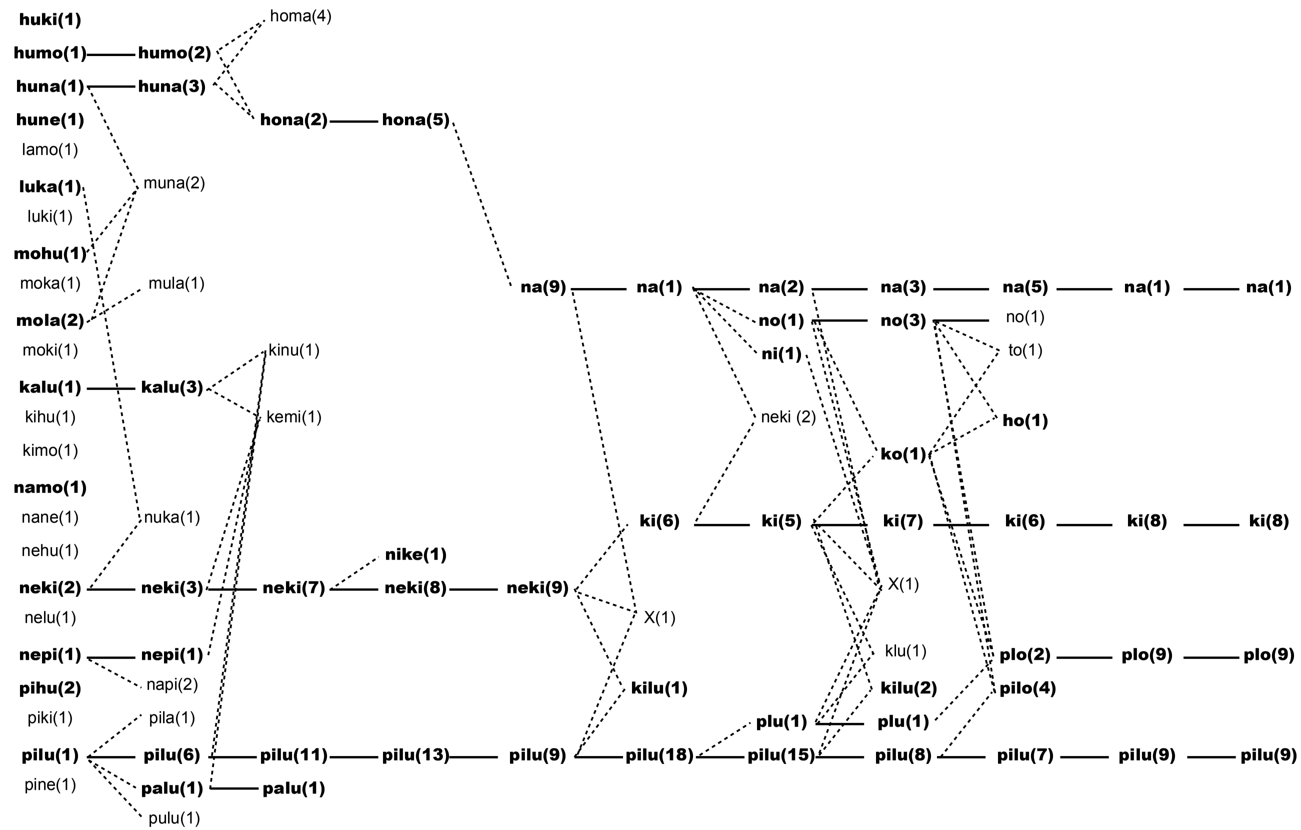


Figure 5.4: A coalescent tree showing relationships of descent between all 27 signal variants in the final segment position over ten generations from a compositional language in the filtered condition. Seen variants are shown in bold, solid lines indicate perfect replication, and dotted lines indicate possible lines of descent with modification. Numbers shown in brackets indicate the frequency with which each variant appeared at, for each generation. This frequency information suggests that the language is adapting to a meaning-space that consists of three elements.

Quantifying the Emergence of Compositionality

Following on from the previous discussion, Cornish *et al.* (2009) present a method for directly quantifying the emergence of compositionality, based on an application of *RegMap* (see: Tamariz & Smith, 2008; Tamariz, *in press*). *RegMap* measures the regularity of the mappings between meanings and signals. It is different however to the Mantel test described in §3.4.3 in that rather than working at the level of whole signals and meanings, it instead looks at the correspondence between signal segments, signal variants, meaning elements, and meaning variants. *RegMap* is:

> "...an information-theoretic metric that combines the conditional entropy of meanings given signals and of signals given meanings and normalises the result to make it comparable across systems of different sizes. Informally what *RegMap* (short for regularity of the mappings) does is return the degree of confidence that a signal element consistently predicts a meaning element (for instance, the degree to which we can be sure that the beginning of the signal encodes color)." (Cornish *et al.*, 2009:196)

Whilst the Mantel test is a general measure of correlation between meanings and signals, it does not differentiate between compositionality and other types of structure that we have found such as underspecification. *RegMap* on the other hand is targeted at measuring this precise kind of structure. It can be formally defined by the following equation (taken from Cornish *et al.*, 2009),

$$(1) \qquad \mathrm{Re}gMap = \sqrt{\left(1 - \frac{H(S \mid M)}{\log(n_s)}\right) \times \left(1 - \frac{H(M \mid S)}{\log(n_m)}\right)}$$

where $H(S \mid M)$ is the conditional entropy of a signal segment given a meaning element (telling us how uncertain we are on average about predicting, for instance, what colour an object is if we hear the first segment of its signal) , $H(M \mid S)$ is the conditional entropy of a meaning element given a signal segment (or how uncertain we are on average about what initial signal we should produce if we know the colour of an object), n_s is the number of different signal variants in the relevant segment position, and n_m is the number of different meaning variants that the particular meaning element can take. By taking the log of these last two values, we can normalise the values between 0 and 1, enabling us to compare across different systems, and by subtracting these normalised conditional entropies from 1, we return the levels of certainty instead of the levels of uncertainty.

Figure 5.5 shows how *RegMap* quantifies the system that emerged in Chain A in the filtered condition. Each graph represents the *RegMap* values calculated for our three meaning elements (motion, shape and colour) in each of our signal segment positions (first, middle, and final). In order to establish statistical significance for these values, a Monte Carlo analysis involving 10,000 randomisations of the possible mappings between meanings and signals were performed (see §3.4.3 for a general description of Monte Carlo tests). The distributions of the values obtained by these randomisations are shown in box-plots: points above (or below) these distributions represent significant divergence from that which we would expect at random (more than two standard deviations away). Fig 5.5.upper shows this information for the subset of the language that is actually given as training input to learners (i.e. after the semantic bottleneck and filtering processes have been applied), whereas Fig 5.4.lower shows this information for the whole language.

If we begin by looking at these lower graphs, we can see that from initial levels that are indistinguishable from random, the *RegMap* values do increase to significant levels for each signal segment. However, this does not happen at the same rate for all segments, but appears to be a gradual process spread out over several generations. First of all motion becomes encoded by the final segment in generation three, then colour is encoded by the first segment in the fifth generation, before finally shape starts to become significantly encoded by the middle segment from the ninth generation onwards.

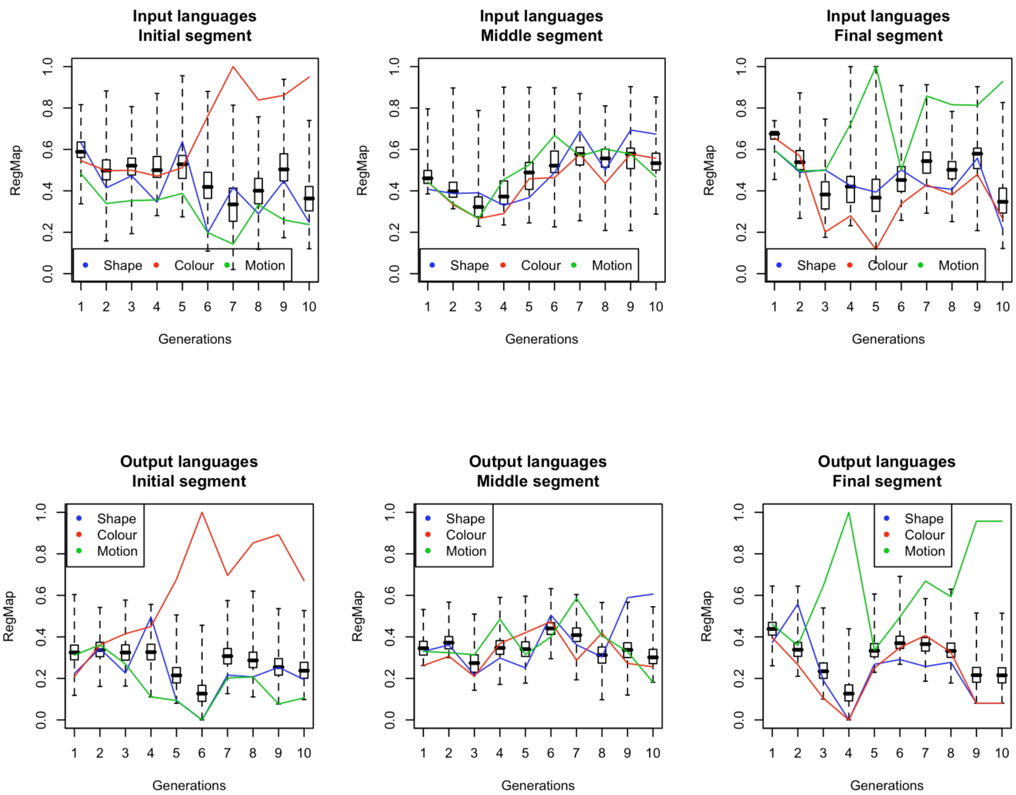


Figure 5.5: The regularity of the mappings between signal and meaning elements changes over time. RegMap quantifies how in a compositional system (Chain A, Exp III), signal segments become consistently associated with distinct meaning elements. Each coloured line represents RegMap values for how reliably a given meaning element (shape, colour, or motion) is being encoded by a particular segment (initial, middle or final). Box-plots show the distributions of these values obtained with 10,000 randomised languages as a control: any points above or below the whiskers therefore show significant differences to what we would expect to see by chance. Upper graphs show RegMap values of the subset of the language used in training for each generation, after the semantic bottleneck and filtering occurred, whereas the lower graphs represent the RegMap values obtained in the language as a whole. Differences between the two therefore reveal the effect that these transmission pressures are exerting on the system. Taken from Cornish, Tamariz & Kirby (2009) with permission.

Competition between Meaning and Signal Variants

These *RegMap* calculations can not only tell us at what point in the history of the language that associations between meanings and signals arise, they can also tell us something about the way in which meaning elements and signal segments compete with one another. In order for a language to be compositional, it is important that the system evolve to avoid ambiguity. If we compare the upper and lower graphs in

Fig 5.5, we can find evidence for competition between meanings all vying for expression by the same signal segment. As an example, Cornish *et al.* (2009) point out that in the input to the third learner, the final segment equally encodes both motion and shape (Fig 5.5.upper). However, rather than reproduce this conflict in their own output, the learner resolves the issue by ignoring the association linking the final segment with shape, and instead amplifies the association with motion, which we see reflected in the *RegMap* scores in Fig 5.5.lower.

Conversely, cases of signals all vying to express the same meaning also occur. If we look to the input to generation 5 we discover that the *RegMap* values for colour are similar in both the initial and the middle segment positions. If we look to the output graphs we find that the learner at this generation resolves this conflict by ignoring the association with the middle segment, and massively amplifying the strength of the association with the first segment only. This shows us that not only are individual signal elements coming to encode specific elements of meaning, but that the system as a whole is adapting to avoid ambiguity.

How Transmission Amplifies Structure

The *RegMap* analysis has so far shown us when signals come to encode meanings, and also that the system resists ambiguous mappings. However, we have not yet discussed precisely *how* these two events come about. What exactly is going on during transmission to make this happen? Making a comparison between the upper and lower graphs does reveal one striking fact, however: in most cases, the absolute values of *RegMap* are slightly higher in the input, as compared to the values in the output. More specifically, the lower the *RegMap* values are in the language as a whole, the more likely they are to be amplified in the subset of the language given as training. What does this mean?

In essence, when we compare the input languages to the output languages, we are seeing the effect that our transmission constraints (the joint action of the semantic bottleneck and the expressivity filter) are having on the system. What the difference between the two graphs is telling us is that when we take a sub-set of the language and filter out homonyms, we are on average finding more structure in that sub-set than we do when we examine the language as a whole. This is potentially counter-intuitive, as it suggests that the less data we encounter, the more regularity we perceive. Particularly early on in the chain when there is very little global structure, a small random sample of meaning-signal pairs taken from the whole language is more likely to contain some regular patterns by chance, than a more comprehensive sample of the language would do. At the very least, it is likely to contain less counter-evidence *against* such patterns existing.

Of course, it could well be the case that a subset of the language genuinely reveals no structural patterns at all. In this case, the language should be transmitted with just as little structure as it had before. The important point to note is that *the illusion of structure* only has to occur once for it to actually become a reality for the next learner. In this way, structural increase is inevitable as long as learning is done on incomplete data. As Cornish *et al.*, (2009: 200) explain:

> "The smaller subsets sampled as inputs to the next generation may locally contain more systematicity than the entire language. Iterating this learning process using these smaller samples therefore provides a platform that allows systematic patterns to be noticed, remembered, and replicated preferentially, thereby allowing them to gradually accumulate in the languages as a whole."

Structural regularities in our languages therefore arise as a consequence of chance patterns that are observed locally being generalised to the language as a whole[18].

[18] It should be noted that this same process probably accounts for the increases in structure in all of the iterated language learning experiments described so far, even when there is no semantic bottleneck or filtering constraints in place (as was the case in Exp II). Even in this instance, we know that there are naturally occurring memory constraints operating that we have already seen effectively doing the same thing as the semantic bottleneck.

5.2.3 Summary

This study has investigated how languages evolve when there is both a semantic bottleneck and a pressure for expressivity present during transmission. Homonyms were filtered out of participants training input in order to encourage signals to uniquely express more of the meaning-space. As with the previous iterated language learning experiments, when the resulting languages were analysed, it was found that transmission error decreased over time as measures of structure actually increased (Fig 5.1). Comparisons between the current study (the filtered condition) with the first study (the unfiltered condition) indicated that our indirect pressure for greater expressivity was working: the number of distinct signals in languages in the filtered condition were much higher, and maintained more stably over time (Fig. 5.2).

Qualitative analysis of the languages themselves revealed that two of the four chains show signs of compositional structure. If we want to learn more about how compositional mappings between individual components of signals and individual components of meanings arise, there are two different methods. The first is indirect; by looking at how signal variants evolve over time in coalescent trees we can detect frequency patterns which suggest when signals start to reflect meaning structures (Fig. 5.4). The evolution of signal forms alone cannot tell us exactly which meaning element a given signal comes to encode. However, the second method can.

Using a measure of structure known as *RegMap* (Tamariz & Smith, 2008; Tamariz, *in press*) we can precisely quantify not only when regular mappings emerge, but also chart the competition between signals and meanings vying to express one another, and also examine the way in which transmission amplifies structure within a language (Fig 5.5). It was argued that much of the structure-generating effects of the transmission process can be explained with reference to the fact that it involves participants only seeing a sub-set of the data: by chance, sub-sets may 'accidentally' contain more structure locally than in the system as a whole. When this data is used as the basis for generating new data, weak structural relationships get amplified, increasing their influence in the future.

5.3 Discussion of Experiments I and III

Experiments I and III differ only in terms of the presence or absence of an expressivity constraint. Nevertheless, with this small change we find the emergence of languages exhibiting some very different structural features. This is interesting in four ways.

Trade-off between learnability and expressivity

Firstly it confirms our intuitive understanding of the tension that might exist between learnability and expressivity. From a purely logical standpoint, it stands to reason that the most learnable systems should also be highly inexpressive. In the extreme example, the simplest kind of language to acquire should be one in which there is just one word for everything (or perhaps, even no words at all). Conversely, more expressive systems should tend to be harder to learn: the more meanings we have to differentiate, the further we get away from the 'ideally learnable' system of one word.

Our experiments have confirmed both of these intuitions empirically. In the unfiltered condition where we have no pressure to be expressive we see the number of distinct strings fall to extremely low levels - to just two signals in chain B. However, whilst these languages may not be able to uniquely express more than a handful of meanings, they are highly learnable and stably transmitted. In the filtered condition on the other hand, where there is a subtle pressure to express more of the meaning-space, we find the opposite: we find highly structured systems that arise and convey a greater proportion of the total meanings, but these systems are not acquired as faithfully. We can therefore see such compositional systems as a trade-off between the twin pressures of learnability and expressivity. This confirms previous simulation results of iterated language learning.

Non-intentional emergence of language-like structure

The second way in which the appearance of such radically different structures is interesting relates to our aim of exploring the potential explanatory power of nonintentional mechanisms of change operating within language evolution. The different outcomes prove that the languages really are created nonintentionally; we know this because the only difference between the two conditions was actually invisible to participants. This meant that participants in each experiment experienced identical learning conditions. The crucial experimental manipulation occurred off-line, meaning that there was no way for participants to be aware of which condition they were in, and therefore no way in which they could have known to alter their behaviour. Our conclusion must be that the differences between the languages in each condition did not arise because participants consciously designed them, or even because they intended to make any change at all. Participants in both conditions are doing exactly the same thing - attempting to replicate the languages exactly as given - yet very different structures emerge as a result.

Cultural Transmission Adds Something to Iterated Learning

Related to this, the third point of interest is that the emergence of qualitatively different solutions to the task tells us that transmission really is capable of adding something to our understanding of *where* structure in language comes from. This may just seem like a restatement of the previous two observations, but it is a point currently worth emphasising because of the results of several recent mathematical models (Griffiths & Kalish, 2005, 2007; Kirby *et al.,* 2007; Ferdinand & Zuidema, 2009). These models are all Bayesian versions of the standard ILM (see §3.2.1), and each present three different (potentially conflicting) accounts of the role of transmission in iterated learning.

Recall that in BILMs, agents' inductive biases (what they bring to the task of learning) are explicitly encoded in the form of a prior distribution over hypotheses (*P(h)*). This distribution dictates how likely the agent is to assume a hypothesis is correct before it has even seen any data. These prior beliefs interact with the data being transmitted according to Bayes rule:

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)}$$

where *P(d | h)* is the likelihood of the data being explainable by a given hypothesis, and *P(d)* is the probability of observing the data averaged over all hypotheses (this acts to normalise the equation)[19]. Using these three pieces of information, the agent can calculate the posterior probability (*P(h | d)*) that a specific hypothesis could generate that data - in other words, the agent can choose which hypothesis to use in order to reproduce that data for the next learner.

In one of the very first applications of Bayesian iterated learning, Griffiths & Kalish (2005) found that the stationary distribution of transition matrix probabilities between all possible languages[20] (in other words, the actual outcome of iterated learning) precisely matched the agents' prior distributions. The implication from this is clear. It suggests that the learner's prior biases alone determine the outcome of iterated learning: cultural transmission adds nothing to the process.

---

[19] These two terms (*P(d | h)* and *P(d)*) quantify the role of cultural transmission in the models.

[20] The stationary distribution of transition probabilities basically tells us the final state that would result from running a simulation for as long as it takes to reach the stable equilibrium where no further change happens. Fortunately there are well understood mathematical techniques for extracting this stationary distribution without having to run each simulation for the (potentially) extremely large number of generations it would take to reach this point.

Whilst this finding does leave some room for debate -- learning biases can of course be acquired via experience with the world and interaction with others, and are not necessarily biologically innate (Griffiths & Kalish, 2007) -- obviously this result suggests that the scenario presented in §2.2.2 and §2.3 which stresses the importance of cultural transmission as an evolutionary mechanism capable of explaining the appearance of design, could be wrong.

In order to investigate this further, Kirby, Dowman & Griffiths (2007) performed another BILM with one difference: rather than agents *sampling* over posterior probabilities of the hypothesis (i.e. having the chances of a hypothesis being selected be proportional to its strength) as in Griffith & Kalish (2005), agents instead selected the *maximum* posterior probability. This change in the hypothesis selection strategy had an immediate effect. Rather than mirroring the prior biases of the learners, cultural transmission was shown to amplify them. Interestingly, they showed that the strength of initial biases had very little effect on the final outcome. Even weak biases got amplified.

Both of these studies only focused on properties of the agents, such as the strength of their initial biases and what strategies they use to select hypotheses. In a third study, Ferdinand & Zuidema (2009) extend this work by exploring social properties such as the size of the population (monadic or polyadic) and bias heterogeneity within agents (whether population consists of agents who all have the same or different biases). They too conclude that cultural transmission adds something to the explanation, and show that for population sizes greater than one, or when there is heterogeneity within the population, even agents who sample their hypothesis from the posterior have stationary distributions that are different to their priors.

The results of Exp I and III adds support to this stance, as they show that an externally imposed cultural transmission constraint can generate a different outcome to iterated learning, despite all the learners possessing the same learning biases. This suggests that we should proceed with caution when using iterated learning as a diagnostic tool for revealing the prior biases of learners, as is done for example in Kalish *et al.* (2007). In this experimental study, participants are given a function learning task to complete. During training, participants were shown a horizontal bar of varied lengths and had to adjust the height of a vertical bar until they were satisfied. At this point they were given feedback on where they should have located the bar, and the next training item appeared. During the test, participants were given the same task, but received no feedback. The data points collected from the test were transmitted to the next learner, until nine generations had completed the task.

The results from this study were shown to reflect known inductive biases of learners: over generations, participants tended to converge on the positive linear function, even when chains were initialised with negative linear functions or just random points. It could well be the case that cultural transmission in this instance is not contributing anything significantly different to the particular learning biases involved, but we cannot be sure. The fact that learning biases can be acquired (presumably culturally) is something that researchers need to carefully untangle when making modelling assumptions or interpreting experimental results[21].

In any case, if we refer only to the data presented in this body of work, the fact that the constraints on cultural transmission were the only thing to change between the two conditions shows us that iterated learning is doing more than just revealing the prior biases of learners.

It Does Not Matter That Participants Already Have Language

The fourth and final way in which the contrasting results of Experiment I and III are interesting concerns a possible criticism that could be levelled at all of the results described so far. Namely that as all of the experiments involve participants who

---

[21] For instance, work by Ferdinand & Zuidema (2008) shows that replicating the Griffiths *et al.* (2007) study with graduate students specialising in mathematics and logic reveal different prior biases. In particular, it seems that experience of working with functions changes a learner's response.

already have a linguistic system in place (perhaps even several linguistic systems), it could feasibly be the case that the structures we see emerging are simply the reflections of the native languages of the learners. If this is the case, what light can studies such as these really shed on our understanding of language evolution?

This is not a criticism to take lightly. Although similar accusations can be levelled at many of the other experimental paradigms that investigate the emergence of novel communication systems (e.g. Galantucci, 2005; Garrod et al, 2007; Healey et al, 2007; Theisen et al, 2009; Scott-Phillips et al, 2009), these studies do at least rely on a different communication medium, either using graphical means or physical movement to convey meanings. As such, they may not tap quite as directly into 'linguistic structure' as studies like these.

Fortunately, there are two reasons for believing that this is not what is going on here. Firstly, our invisible modification appears to shape the properties of the emergent languages much more than any similarity to the languages of participants. Secondly, we find that these experimental results are backed up by the computational models already described. Agents in these models have no prior linguistic system in place or any language specific learning biases in place, but nevertheless go on to develop systems with the kinds of properties found here. The most parsimonious explanation is therefore that the structures we have seen appearing in the experiments arise from transmission constraints and the adaptive process of iterated learning, rather than being the product of underlying native language competences (Kirby *et al.*, 2008a; Cornish *et al.*, 2009).

There is a small caveat to this, as Cornish *et al.* (2009: 201) note:

> "We fully expect that language evolution through iterated learning will involve adaptation to all aspects of the transmission bottleneck, and this will include the biases of language learners... [P]articipants bring to bear a mixture of biologically basic biases and those that arise from their acquired cultural heritage. We can

> see no principled way to separate these out. This means that our experiments should not be taken as a 'discovery procedure' for uncovering our evolutionary ancient learning biases but rather as a tool for understanding the fundamental adaptive dynamics of the cultural evolution of language by iterated learning."

This means that although we might expect to find that the native languages of the learners do interact in the process at some level (for instance, certain forms might appear to be more salient than others based on similarities or consistency with existing language structures, and thus will be preferentially retained by learners), this does not mean that these interactions alone are responsible for the appearance of design. There are many biases at work, and transmission is key to understanding how these biases manifest themselves.

5.4 Experiment IV: Increasing Transmission Fidelity

The previous experiment has shown us that having a pressure for expressivity is a necessary requirement for compositional structures to emerge. Nevertheless, it does not seem to guarantee compositionality. It only appeared to emerge in two of our four chains, and even when it did appear, it was not stably transmitted to future generations. Perhaps one of the reasons for this is the fact that participants' levels of recall are actually quite low throughout.

As an example, if we take a look at the normalised Levenshtein Distance scores we have obtained from the first generation of learners in all three experiments so far (Fig 4.1.right, Fig 4.4.right, and Fig 5.1.right), we find that participants are struggling to accurately learn the items they are being exposed to during training. Of course the languages given to the first generation represent the hardest learnability challenge of all[22]. Nevertheless, it is striking that the measure of transmission error shows learners are only getting between 20-35 per cent of signals correct. This is

[22] Due to the fact that all three studies investigate different experimental conditions, the first generation represents the only point at which we can safely aggregate performances.

hardly surprising when we consider that they only see each training item six times in total. If the amount of exposure to each training item was increased, would it lead to the emergence of compositional languages that were more stable? This is the question that Experiment IV seeks to address.

5.3.1 Method

Aims and Experimental Hypotheses

This experiment investigates the question of how important early transmission fidelity is. In order to investigate this, the number of exposures to training items was doubled from six, to twelve. There was both a semantic bottleneck, and an expressivity requirement in place, making it comparable to the previous experiment. Just like the previous experiment, it was hypothesised that learnability and structure scores would increase over time, and that compositionality would emerge due to pressure arising from the filtering constraint. It was additionally hypothesised that the increase in fidelity would help compositionality to be maintained over time. These hypotheses are reproduced below.

1. **The Learnability Hypothesis**: Languages will become easier to learn as a result of iterated learning.
2. **The Structure-Increase Hypothesis:** Languages will become more structured as a result of iterated learning.
3. **The Compositionality Hypothesis:** Pressure to generalise to novel stimuli combined with a pressure to uniquely express each meaning will result in languages evolving to become compositional.
4. **The Stability Hypothesis:** Increasing the fidelity of transmission will result in stable compositional languages.

Experimental Design

Four chains were initialised with randomly constructed languages, containing 27 meanings paired with 27 signals. Participants were trained on approximately half of the meaning-signal pairs, which were presented one at a time, in random order, twelve times each. An expressivity filter was also applied to ensure that every signal was unique. Training occurred over three distinct rounds, with each training item getting four exposures each time. After each training round, participants were given a practise test where they were shown a picture of a meaning, and asked to provide the correct signal. At the end of the final training round, this test was extended to include each of the 27 meanings, including the ones that participants had not seen in training. The output from this final test was collected and used to form the training input to the next generation. The experiment was run using E-Prime software, and the results were analysed using *R*.

Participants

A group of 40 participants were recruited via an advertisement in the University of Edinburgh's student employment service. Of these, 17 were male and 23 were female, with an average age of 21.48 (*S.D* = 3.64). Participants were offered £5 to take part, and in order to be eligible had to meet the following requirements: be fluent in English, not be dyslexic, have normal or corrected-to-normal vision, have not taken part in any previous studies, and not have have formally studied linguistics. Each participant was assigned to one of the four transmission chains at random. The study conformed to the ethics guidelines set by the University of Edinburgh's College of Humanities and Social Science, and participants were fully briefed before taking part of their rights of withdrawal and anonymity.

5.3.2 Results of Experiment IV

Learnability and Structure Increase

In order to investigate the first two hypotheses, the structure and error scores were calculated for each generation of each chain. These values are shown in Figure 5.6. As we would expect, the error and structure scores in the first generation as compared to the last generation show significant differences. Paired one-tailed t-tests indicate a mean decrease in error of 0.35 ($t(3) = 11.079$, $P < 0.0008$) and a mean increase in structure of 8.45 ($t(3) = 5.767$, $P < 0.005$). Examining Fig 5.5 in more detail, it appears that all languages are significantly structured after the third generation, and at least one chain is stably transmitted for several generations. This would seem to confirm hypotheses one and two.

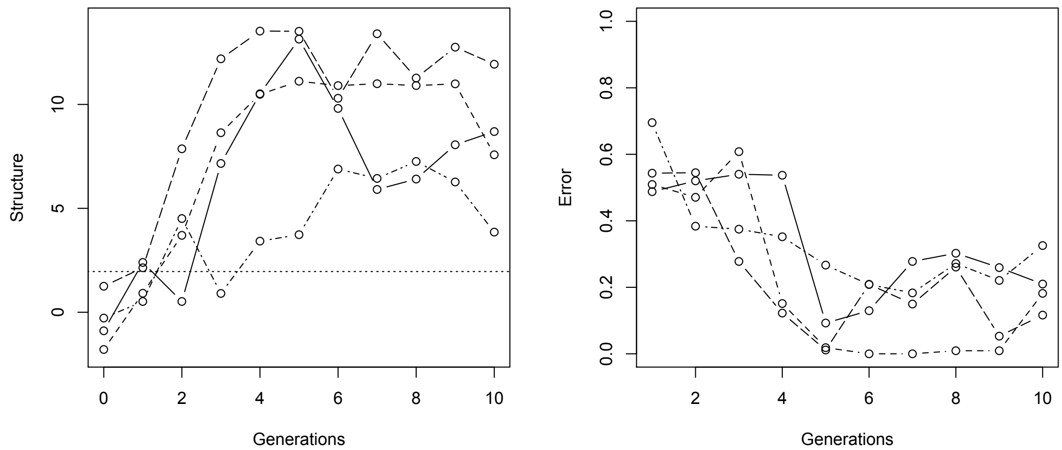


Figure 5.6: Structure and transmission error scores over generations for chains in Exp IV. These graphs show that the languages are becoming easier to learn, and more structured over time. Additionally, it appears that at least one chain is being transmitted faithfully over multiple generations. Transmission error (right) is measured using normalised Levenshtein Distance, whereas structure (left) is calculated using z-scores derived from a Monte-Carlo analysis of the correlation of the distances between meanings and signals. Points above the dotted line (left) indicate significant structural regularities between meaning-signal mappings. Re-drawn from Cornish (2010) using 10,000 Monte-Carlo randomisations instead of 1,000.

Transmission Fidelity Increases

Before we continue, we must first examine whether or not increasing the amount of training has had the desired effect of increasing the fidelity of transmission. In other words, has error decreased as a result of participants receiving more exposure to training items? To test this, we can compare the performance of learners on the initial randomly generated languages in both the single (Exp III) and double (Exp IV) training conditions. Although we predict that these languages are harder to learn on average than later ones, we cannot be sure that languages at later generations are comparable as they are likely to be differently structured. These transmission error scores are shown in Table 5.2 below.

Exp III (single)	*Exp IV (double)*
0.7407407	0.4876543
0.7913139	0.50925926
0.7098765	0.6953263
0.6406526	0.54320988

Table 5.2: Transmission error scores found at generation one in the single and double training conditions. These scores indicate that increasing the amount of training is indeed increasing the fidelity of transmission.

If we run an unpaired one-tailed t-test on these scores we find that the difference between them is significant ($t(3) = 2.8633$, $P < 0.01$): recall in the double training condition was significantly higher (mean error 0.558) than in the single training condition (mean error 0.721).

Compositionality is Not Stable

Our third hypothesis predicts that we will find compositional languages appearing in the chains as a result of the trade-off between learnability and expressivity pressures. Examining the raw data (Appendix B4), it appears that we find at least one instance of compositionality, arising at generation four of chain D. This system is reproduced in Table 5.3.

wakimo	hunkimo	pokimo	□
wakemo	hunkemo	pokemo	○
waknimo	hunimo	ponimo	△
wakiki	hunkeki	pokeki	□
wakeki	hunkiki	pokiki	○
wanikuko	hunikuko	ponikuko	△
wakikuko	hunkikuko	pokikuko	□
wakekuko	hunkekuko	pokekuko	○
wanikuki	hunikuki	ponikuki	△

Table 5.3: An example of a compositional system arising at generation 4 in Chain D in the double training condition. This language has 27 distinct signals, and has a very similar structure to the system which emerged in the single training condition (Exp III). Signals in this system are composed of three morphemes: colour-shape-motion. Redrawn from Cornish (2010) with permission.

This system bears a striking resemblance to the compositional language that emerged in the previous experiment. Every meaning has a unique signal, and there is a clear pattern to how signals are internally structured. The first segment represents colour: **wa-** for black; **hu-** for blue; and **po-** for red. The second segment represents the shape of the object: **-ki-** for square; **-ke-** for circle; and **-ni-** for triangle. The final segment represents the motion of the object: **-mo** for horizontal; **-ki** for bouncing; and **-kuko** for spiralling. There are some minor deviations from these general rules, but even these appear to only apply in a predictable context. For instance, bouncing triangles acquire the same suffix as spiralling squares and circles, whilst spiralling triangles acquire a suffix which appears to be a combination of the standard bouncing and spiralling suffixes.

It is good to see another strong example of a compositional system arising in a chain that has an expressivity filter applying. However, the main aim of the double training condition was to see whether this encourages any compositional system that might arise to be more stably maintained. Has this occurred here? Examining the language at generation 10 of this chain reveals that this system does not last. Not all elements of this system have been lost, however. In fact, looking a little closer, we find something curious. Table 5.4 reproduces this language.

wanimo	henimo	ponimo	□
wanimo	henimo	ponimo	○
wanimo	hekiko	ponimo	△
wakiko	hekiko	pokiko	□
wakiko	pokiko	pokiko	○
wakiko	hekiko	pokiko	△
wahikeko	hehikeko	pohikeko	□
wahikeko	hehikeko	pohikeko	○
wahikeko	hehikeko	pohikeko	△

Table 5.4: An example of a language at generation 10 that was previously compositional and is now underspecified. What is surprising about this structure is that it arises in a chain (D) involving both filtering and double training. Parts of the signals correspond to colour and motion features, however, shape is no longer explicitly encoded.

In spite of the fact that the languages are still being filtered for homonyms at every generation, we nevertheless find that this system is underspecified. Signals can still be decomposed into compositional parts however, but instead of differentiating all three meaning elements like before, now signals are composed of just two parts: colour and motion. Examining the signal variants themselves , we find little change in those used to refer to colour. We find **wa-** still corresponds to black, **hu-** has changed to **he-** to describe blue objects, and **po-** is still used to describe red items.

However, it appears that the variant which previously meant triangle has merged with the old suffix for horizontal to form **-nimo** (indicating horizontal movement); there is a new suffix **-kiko** to describe bouncing items; and a new suffix **-hikeko** to describe spiralling items. Both of these latter suffixes are possibly related to the old motion suffixes combined with one of the old shape variants.

Compositional Underspecification: A Stable Compromise?

The mixed system shown in Table 5.4 is not a one-off. Recall that our examination of Fig 5.5 revealed that at least one of the chains resulted in a system that was stably transmitted for at least five generations in a row. An examination of the structure of this languages, taken at the mid-point of its stable run (generation 7) reveals that it too exhibits properties of both a compositional and underspecified system. It is shown below in Table 5.5. This system again settles upon a solution to the problem of transmission by only encoding the colour and motion features of the meaning-space. Again, colour is represented by the first segment: **[null]** in the case of black items, **pa-** in the case of blue items, and **me-** in the case of red items. Suffixes encode motion: **-linu** for horizontal, **-gahili** for bouncing items, and **-wenu** for spiralling items.

linu	palinu	melinu	□
linu	palinu	melinu	○
linu	palinu	melinu	△
gahili	pagahili	megahili	□
gahili	pagahili	megahili	○
gahili	pagahili	megahili	△
wenu	pawenu	mewenu	□
wenu	pawenu	mewenu	○
wenu	pawenu	mewenu	△

Table 5.5: Another example of a language that is both underspecified and compositional arising in a chain involving both filtering and double training. This example comes from Chain B, generation 7. This language was stable for five generation.

The systems that emerge in Tables 5.4 and 5.5 are remarkable in that participants in these conditions never see any examples of homonyms in their training data. In order for underspecification to survive in languages that are being filtered, a delicate balancing act must be maintained. As long as homonyms are evenly distributed throughout the language (as they are here), and there is some degree of compositionality that allows participants the chance to reconstruct any form unlucky enough to not make it through the transmission bottleneck, the filtering process can be bypassed. This kind of system therefore represents an elegant, albeit unexpected, solution to the particular transmission constraints being applied here. The fact that it is transmitted more faithfully than any other that we have encountered so far is a testament to the fact that is is also highly adaptive.

5.3.3 Summary

This study examined whether increasing the amount of exposure participants had to training items would increase the fidelity of transmission, and lead to the emergence of compositional systems that were stable. Four transmission chains were run with a semantic bottleneck and filtering, to encourage the emergence of compositional structure. In contrast to Exp III, participants also received double the training. A comparison of the levels of recall by the first generation of learners in the single and double training conditions revealed that the extra training was helping participants to acquire the signal-meaning pairs more faithfully. Qualitative analysis of the resulting languages however showed that only one instance of compositionality was recorded, and that this was not stably transmitted to future generations.

Instead, that particular system changed to incorporate features of both compositionality and underspecification. An almost identical system to this one was also found in another chain. In this particular case, that system *was* stably transmitted, over five generations. The fact that an underspecified system could emerge in a condition where homonyms were filtered out before transmission might at first appear surprising, but reflects the fact that there always was, in a sense, an optimal solution to bypassing the expressivity filter. Although this result was certainly not anticipated by the author in advance (and, I would argue, could not have been engineered consciously by participants, even if full disclosure was given of the fact that their data was being culturally transmitted to others), it is a reminder that cultural evolution is capable of adapting in surprising and unpredictable ways.

5.5 Discussion of Experiments III and IV

The fact that we did not find stable compositional languages in Experiment IV definitely adds weight to the idea that compositional systems are fundamentally harder to acquire than underspecified systems. Even though a perfectly compositional system was created early on in one of the chains, the increased training was still not enough for it to be faithfully acquired by later generations.

How then can we account for the fact that in computational simulations of iterated language learning, compositionality is not only consistently found in all runs, but also highly stable when it does emerge? Perhaps the most significant difference between the ILMs and these transmission chain experiments lie in the number of generations of learners that they employ. Simulations of artificial agents are typically run for hundreds if not thousands of generations before stable compositional systems emerge. Therefore one explanation for our lack of success here is that perhaps the transmission chains need to be allowed to run for longer.

The vast number of generations required in some ILMs has actually been used as a criticism against the ecological validity of such models. For instance de Beule & Bergen (2006) point to a number of studies which show that pidgins and creoles emerge with compositional languages in just a few generations, making the models appear unrealistic in comparison. The results of the studies presented here supports this idea that human learners are much faster at converging upon compositional systems - in both experiments III and IV we find cases of compositionality arising in as few as four generations. A more obvious difference between the cases of evolving pidgins and creoles, and the experiments here concerns the structure of the populations involved. One avenue of work which is currently being explored at the LEC in Edinburgh involves increasing the number of learners per generation. Early work suggests even adding just one more learner to each generation can result in the emergence of more stable compositional systems (Line, 2010; Winters, 2009).

Going back to the languages that we did find in the double training condition, the fact that two of them seemed to find the perfect structural balancing act between the two system types (compositional and underspecification) really is quite remarkable. Consider again that in spite of the fact that participants never see duplicate signals in their input, they still end up being in perfect accord with previous generations in where to posit duplicate signals in their output. This was not a solution that any of the participants (or even the experimenter for that matter) could have anticipated in advance, and yet the nonintentional processes of cultural transmission delivered it.

Chapter Six

Language Adapts to Sequence Learning Biases

The picture we have built of language so far is that over the course of its transmission from learner to learner, it encounters many different constraints which can (over time) impact upon its structure. Each of these constraints is a different kind of bias that the language is adapting to. Some of these biases are internal to the learner, and others are external. For instance, in chapter 4 we explored a kind of bias that was externally imposed - the semantic bottleneck - and contrasted it to the naturally occurring phenomenon of imperfect learning. In chapter 5 we looked at another externally imposed constraint in the form of the filtering process that removed homonyms from the input to learners and encouraged greater expressivity. The fact that the presence of these external biases can be shown to have such major impacts upon the resulting systems is a reminder that we must be careful when making claims that we can use iterated learning as a means to uncover those biases that are internal to learners. As Smith *et al.* (2008: 534) point out:

> "mental properties cannot simply be read off from [properties of language], because the cultural process mediating between aspects of the mind and features of language distorts the underlying biases of human learners."

That being said, it is clear that at least some, or indeed the majority, of the kinds of adaptations that we have seen the languages in these experiments undergo clearly have been the result of learning and processing mechanisms that are internal to our human participants. Our most consistent finding is that languages change in ways that make them easier to learn by future learners. Whilst in some sense we have already seen the outcome of such cognitive constraints in the studies presented, their effects have been intertwined with those of many other biases. It would be nice if we could study these mechanisms and learning processes in isolation somehow.

One obstacle to achieving that goal with the current experimental framework is the presence of structured meanings. The meanings to be conveyed represent yet another external constraint to which language must adapt. Consider how in every experiment so far we have seen that the structure of signals comes to reflect the structure of the meanings in some systematic way - whether that involves the meanings themselves undergoing some kind of levelling of features leading to fewer distinct signals being required to express them, or signals becoming decomposable into segments which get mapped onto individual meaning elements. In fact, it could well be argued that given the presence of our fixed and easily decomposable meanings, we perhaps should not be so surprised that we get structured signals out; in a very real sense, it is meanings and our need to differentiate them which cause this structure to appear.

Given that we find ourselves in a position where the previous work has focused exclusively on the cultural transmission of meanings and signals - or rather, how the presence of structured meanings can give rise to language-like structures - we can ask ourselves another question. Can other types of cognitive constraints, in the *absence* of meanings, give rise to any interesting structural features? In other words, what happens when the only things being culturally transmitted via iterated learning are signals? This final experiment explores just this, by attempting to isolate the effect of sequence memory constraints on cultural transmission.

6.1 Sequential Learning Constraints

Language, either spoken or signed, consists of a complex arrangement of signals that can be described in terms of statistical relations between different units (Conway *et al.*, 2007). These signals are necessarily organised sequentially. This stems from the fact that the communication channel itself demands that transmission be serial. As such, we expect that the ability to encode and manipulate sequential patterns should be an important pre-requisite for using language (Lashley, 1951). This is indeed the case. Not only is there is a strong link between sequence memory, and both word learning and vocabulary development (Gupta & MacWhinney, 1997; Baddeley, 2003), but a number of psychological studies have also linked deficits in sequence learning with a range of different language disorders (Plante *et al.*, 2002; Hoen *et al.*, 2003; Christiansen *et al.*, 2010). At the same time, artificial language learning (ALL) studies have shown that sequential learning is implicated in many aspects of normal language acquisition, from segmenting speech sounds (Saffran et al., 1996), to detecting long-distance dependencies between different words (Gomez, 2002; Onnis et al., 2003).

Taking this into consideration, it has been suggested that language has evolved to fit these sequential learning and processing mechanisms in the brain (Christiansen, 1994; Christiansen & Chater, 2008). This approach stresses the fact that these cognitive mechanisms originally evolved for purposes other than language. Although sequence memory and sequential learning abilities are employed extensively in language, they are in fact domain-independent mechanisms, involved in motor control and planning, as well as working memory (Lashley, 1951; Christiansen & Ellefsen, 2002; Baddeley, 2007). We have already seen empirical investigations of this idea that universal properties of linguistic structure can be explained by these non-linguistic constraints on learning. For instance, Christiansen & Devlin (1997) show how word order patterns that match observed typological distributions in the real world can be derived from models relying on very general sequential learning mechanisms, whilst Ellefsen & Christiansen (2000) show how linguistic subjacency constraints could be derived from limitations on sequential learning using both connectionist models and ALL studies involving humans (see §3.2.3).

Whilst studies like these indicate that there is an apparent fit between universal properties of natural language structure and these general cognitive constraints, this relationship has only been shown in humans indirectly, via tests of comprehension. In other words, participants have been tested on different types of structure created by the experimenter and shown to only be able to acquire those structural patterns that are in some sense 'naturally' occurring (Ellefsen & Christiansen, 2000). When combined with computer simulations that replicate this same behaviour, this strongly implies that there is nothing specific to language about the mechanisms responsible. Nevertheless, the argument would be strengthened if we could not only observe these structural patterns being easily acquired by individuals in an ALL experiment, but also to witness them actually emerging culturally in a population of learners, from a starting point of no structure.

In the next study we will investigate whether these biases in participants' ability to process and recall sequences can lead to the cultural evolution of structure. Importantly, we will change our framework slightly to try to remove any other biases that might be acting upon the signals, and we will initiate the transmission chains with signals that do not contain any structural regularity.

6.2 Experiment V: Transmitting Signals With No Meanings

6.2.1 Method

Aims and Experimental Design

In all of the previous experiments we have defined a language as a set of mappings between signals and meanings. The aim of this experiment is to investigate how sequence memory constraints affect the structure of signals when they are culturally transmitted *without* any meanings[1]. In short, we want to ascertain whether structural regularities appear when there is no externally imposed structure encouraging adaptation. Our working hypotheses will therefore be familiar: we are expecting signals to become easier to learn, and more structured over time.

1. **The Learnability Hypothesis:** Signal-strings will become easier to learn as a result of iterated learning.

2. **The Structure-Increase Hypothesis:** Signal-strings will become more structured as a result of iterated learning.

As there are no computational simulations of this particular experiment, we do not have any firm predictions about the precise nature of the structure that we might find.

Experimental Design

In order to address the hypotheses, a series of eight transmission chain experiments were run[2]. Participants were trained on a set of 15 signals via an implicit learning technique. During training participants were shown each signal very briefly on screen, before being given the opportunity to reproduce the signal they had just seen by typing it out. This only happened after a delay, which forced participants to

---

[1] The design of this experiment benefitted enormously from collaborative discussions held with Morten H. Christiansen, who offered useful advice on constructing the initial languages, implicit learning techniques, and also suggested quantitative methods to analyse the results.

[2] This was done using a custom-built experimental platform created by Simon Kirby using *Processing*.

keep the item in their mind for a short while. Once the signal had been reproduced by the learner, the next signal appeared automatically. After six passes over all of the training data, participants were asked to recall all 15 items. They were not told in advance that they would be required to do this, and no feedback on their performance was offered until the very end of the experiment, when they were told how many items they had reproduced perfectly. Every unique signal that a participant entered during this recall round was accepted, regardless of whether it was correct or not. If participants entered a string that they had already submitted, they were notified of this, and asked to try again. The experiment continued until either all 15 strings had been provided, or participants withdrew themselves[3]. The data collected in the recall round became the new training data for the next learner, and the process iterated until ten 'generations' had passed.

Materials

The design of the initial stimuli was very carefully controlled in this study. Each "language" consisted of a set of fifteen letter-strings. These were initially constructed to ensure that the frequency of bigrams and unigrams was as uniform as possible. Each string-set was composed of six characters (**a-f**), each of which appeared exactly 10 times. The lengths of strings were also controlled so that some were not more frequent than others. Five of the strings were three characters in length, five were four characters in length, and five were five characters in length. Particular attention was paid to the beginnings and ends of strings: each character could only start or end a string a maximum of three times in order to ensure that the distribution of characters did not favour certain start or end sequences. For instance,

[3] In total, three participants excused themselves from the final recall round as they felt they could not remember any of the items. These participants were fully debriefed and received payment like any other participant, but as they did not provide a full set of signal-strings for the next participant, their results were discarded.

if #**a** appeared three times, **a**# could only appear twice[4]. This meant that every initial language had the same approximate distribution of bigrams. This structure is shown in Figure 6.1 below. An example of one of the initial languages is also given in Table 6.1. This shows the language in its underlying form, and not the form given to participants. This distinction will be explained in the next section.

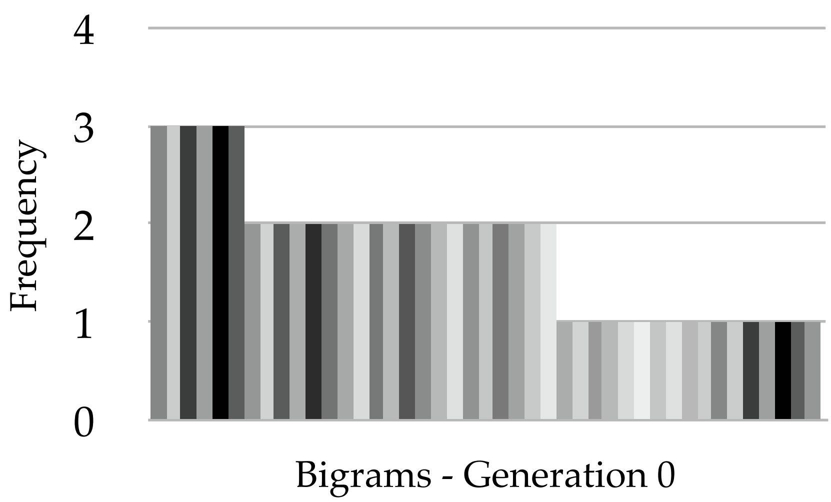


Figure 6.1: The frequency distribution of bigrams in the initial languages of Exp V. Each bar represents a particular bigram (for example 'ac', '#b' or 'ff') and the frequency that that particular bigram appeared in the data. As the figure shows, none of the bigrams appeared more than three times in total, and this only occurred at edge positions. The strings in each language were tightly constrained to have as flat a distribution as possible and avoid any patterns being present from the outset.

[4] The # symbol represents the beginning or end of a string. As these regions (known as the 'anchor points') are particularly salient for participants, extra effort was put into ensuring that these positions were not under or over represented by certain characters (Morten Christiansen, *personal communication*).

initial language
abd
bdc
cae
def
ecd
fcdb
adba
bfde
ceaa
dfed
efdac
fbfcf
aabee
bccfb
cebaf

Table 6.1: An example of one of the initial languages generated for experiment V. This shows the underlying form of the language presented to learners in chain B. These languages were generated with strict controls to ensure that they contained as little regularity or repeating patterns as possible.

Controlling Other Biases

Designing the initial languages in this way ensured that there were no strong regularities in the initial input to learners that could bias them in any particular direction. However, this is not the only type of bias we could imagine operating in this kind of study. Of additional concern was the fact that certain combinations of letters are easier to pronounce than others. Consider the example of **PODDA** versus **LWRRT.** Clearly the former would be much easier to remember than the latter. In order to address this we only used consonant letters to form strings.

There are two other kinds of bias that could potentially affect the learnability of strings, even of those composed entirely of consonant characters: the first concerns the possible introduction of acronyms into the data, while the second concerns the possible effect of the keyboard layout encouraging the emergence of certain typing patterns. In order to avoid both of these biases, at the end of every experiment the languages were remapped onto new consonant characters, and the output of this remapping was visually inspected for acronyms before being given to the next learner. This remapping preserved the underlying structure of the languages, but destroyed the surface structures which might have been influenced by participants' additions. Table 6.2 illustrates what this process looks like using a small example.

In this table we can see that the underlying (plain-text) form of the language in the first generation consisted of three strings **<abc, ddbe, abdde>**. These characters were mapped onto consonant letters, creating the training input given to the first learner: **VTG, DDTR** and **VTDDR**. After training, the first learner attempts to reproduce the strings that they were given. This reproduction is not perfect however, and two new strings have been created. The second one is of particular concern - **DVDTV**. Being as though this string contains salient acronyms, if we gave it to the next learner and it was successfully reproduced we could not be sure whether it was because there was something adaptive about its underlying structure, or whether it was just easily identified and remembered because of the associations caused by the acronym. In order to overcome this, we first 'decode' each string to reveal its plain-text identity, and then apply our remapping procedure again to create a new set of training data for generation two.

GENERATION 1	*GENERATION 1 INPUT*	*GENERATION 1 OUTPUT*	*GENERATION 2*	*GENERATION 2 INPUT*
abc	**VTG**	**VTG**	**abc**	**LHP**
ddbe	**DDTR**	**DVDTV**	**dadba**	**FLFHL**
abdde	**VTDDR**	**VTTGK**	**abbcf**	**LHHPX**

Table 6.2: An example of the remapping procedure used in Experiment V to remove typing biases and the generation of acronyms. The underlying structure is maintained in plain-text form, visible only to the experimenter. The forms that participants actually see consist of capitalised consonant characters. After every generation, these strings are decoded back into plain-text, and then re-encoded and checked for acronyms. If acronyms are present after the remapping, it is repeated with another set of characters until a solution is found.

To recap, the characters **a-f** were only used by the experimenter to keep track of the underlying structure, and allow comparisons over generations. Participants actually received sequences like **VTG, DDTR,** and **VTDDR.**

Participants

In total, 80 participants were recruited for the study, with the vast majority responding to an advertisement placed in the University of Edinburgh's student employment service. Of this number 51 were female and 29 were male (age: *M* = 21.72; *S.D* = 4.08). For this study it was decided that all participants should be monolingual speakers of English. In addition to this requirement, participants were only eligible if they had normal or corrected-to-normal vision, were not dyslexic, and had not taken part in any of the previous experiments. As the experiment lasted less than 15 minutes, participants received £2 and were offered a biscuit for taking part. This study met the ethical guidelines set by the University of Edinburgh's College of Humanities and Social Science.

Procedure

Participants were given both verbal and written instructions about the format of the experiment (See Appendix C). At no point was the experiment referred to as a language task: participants were told that the experiment was exploring their recall abilities, but they were not told how many letter strings there were, nor were they informed about the recall test at the end. Training was conducted using an implicit learning technique. Strings were selected at random and appeared one at a time on the screen for exactly 1000ms, before disappearing. At this point, there was a 3000ms delay where participants could not use the keyboard. If participants attempted to start typing before the 3000ms wait period was over, their string would not appear on the screen and a beep would sound letting them know that they had typed too soon. This delay was included to ensure that participants were implicitly forced to commit the string to short-term memory. After the delay, participants were prompted to reproduce the string that they had just seen by the appearance of a flashing cursor. This process repeated as soon as the learner had entered their response, and continued until each of the 15 strings had been seen exactly six times. On average, this took a little over six minutes.

Once this training phase had been completed, more instructions appeared on the screen. Participants were now told that they had seen 15 different letter strings, and that they needed to try to reproduce all 15 of them as best they could. This part of the experiment was self-paced. A counter at the top of the screen let participants know how many guesses they had left, but they were not given any feedback as to whether their answers were right or wrong. If the same string was entered twice, an error message appeared to let the participant know, and encourage them to make another attempt. Once the final string had been entered, participants were immediately given their absolute score, were given a quick questionnaire to fill in detailing whether they noticed any patterns in the languages or not, and were then fully debriefed about the purpose of the experiment.

6.2.2 Measuring Structure and Learnability

In order to test our hypotheses, we need to be able to measure both structure and learnability in our languages. Recall that in §3.4.3 we defined language as a mapping between meanings and signals, and stated that a language was structured if similar signals get mapped onto similar meanings. Since we have removed meanings from our definition of language (a language is now just a set of 15 different letter-strings), our definition of structure must also change. Signals in our languages have no external referents, so there can be no relationship between parts of the signal and any units of semantic or propositional content. However, that does not mean that signals cannot be composed of parts.

Here we can make a useful comparison to the way some animal communication systems work. For instance, the songs of some species of birds and cetaceans also convey no propositional content, yet nevertheless appear structured hierarchically, involving the re-use of smaller units to form larger units that get repeated (Payne & McVay, 1971; Nelson, 1973; Doupe & Kuhl, 1999; Rendell & Whitehead, 2005; Hurford, 2011). For instance, Brenowitz (1997) describes the hierarchical structure of birdsong in the following way: *notes* (or elements) combine to form *syllables*, which link together to form *phrases* (or motifs), which together constitute a given song *type*. Similar observations have been made concerning humpback whales (Payne & McVay, 1971), whose songs are generally composed of more than a dozen complex units organised into *phrases*, which get repeated to form a *theme*, which gets combined with other themes to form the *song*. Clearly, the structure of signals in these animals is combinatorial but not compositional. There is a syntax, even if there is no semantics.

We can use this insight then to help us describe what we expect the emergence of structure in our culturally transmitted signals to look like. In short, *a language is structured if it contains reusable units which can be combined to form larger units.* This is a fairly broad definition of structure, but it is good as long as we have a reliable way to measure whether signals contain units that are being reused. Fortunately, in the literature on artificial language learning, there is at least one technique that we can co-opt to do this: associative chunk strength (ACS) (Knowlton & Squire, 1994). According to Pothos & Bailey (2000:851):

> "To compute the global associative chunk strength...we considered all the chunks that make up a given test item (i.e., all pairs or triplets of sequential symbols). The associative chunk strength of each chunk is defined as the number of times it appears in the training items. The chunk strength of a test item is calculated by averaging the associative strengths of all chunks in the item."

In other words, it tells us how often on average each chunk (i.e bigram or trigram) that makes up a given signal appears in the training data. We can illustrate it using an example. Let us imagine that a learner is given the following training items to learn from: **abc, abd,** and **abcdef**. Now let us imagine that during the test the learner produces the sequence **abc**. This signal is composed of three different chunks: **ab, bc,** and **abc.** Examining the training input, we see that **ab** occurs three times, **bc** occurs twice, and **abc** also occurs twice. We calculate the ACS for this signal by adding these frequencies together (7) and dividing by the total number of chunks (3). This reveals that each chunk in this signal appeared on average 2.33 times in the training data. We can then calculate the average associative chunk strength (referred to as the global chunk strength) of all signals in our language by adding up the ACS score for each individual signal, and dividing this by fifteen.

If global ACS is shown to increase over time, it means that fragments are being identified and re-used more often. This indicates two things: firstly that learners are grouping together individual characters in order to create these chunks, and secondly, that these chunks are being reproduced as independent units. In terms of determining whether those fragments themselves are going on and combining to form larger units above this level, this will have to be determined via qualitative analysis of the strings themselves[5]. Another way to think about ACS is to imagine its effect on the learner: in languages where ACS is high it makes items appear more familiar, independently of whether those items have been seen before or not[6]. This is because learners are effectively seeing 'bits' of signals that they have not seen, repeated or packaged up in the structures of signals that they have. Although we will not be using any kind of semantic bottleneck here and learners will see all of the data, increased global ACS should help facilitate acquisition in much the same way

[5] There are automatic ways to extract this information. For instance, Suzuki *et al.,* (2006) managed to create an automated classifier system for analysing humpback whale song which could differentiate units and phrases using measurements of entropy and transitional probabilities.

[6] Morten Christiansen (*personal communication*).

that compositionality does -- by increasing the number of tokens that can be generalised from.

It is worth noting that we are using ACS in a slightly unusual way here compared to the literature. Typically ACS is measured to ensure that novel grammatical and ungrammatical test stimuli given to participants has the same distributional structure. This is because items with high chunk strength tend to get rated by participants as being more grammatical (Knowlton & Squire, 1994). In other words, researchers want to be sure that participants are correctly identifying grammatical structures because of rules they have acquired through training, and not because there are noticeable structural differences between the two sets of stimuli. As such, ACS *is* a known proxy for indicating the amount of structure in test items (Pothos & Bailey, 2000). However, it is usually a factor to be controlled when generating stimuli, and not a dependent variable to be analysed over the course of experimentation.

If we move on to how we determine learnability however, we find that we do not need to redefine anything. Like in our previous studies, *a language is learnable to the extent that it is transmitted faithfully without error.* However, we do encounter one problem when it comes to calculating transmission error. In the previous studies we had a way to pair signals from one generation with signals from the other; we used the meanings as a stable link. When we calculated the normalised Levenshtein Distance (nLD) (§3.4.3), we were effectively asking: "how similar was the signal used by learner 3 to describe a bouncing black triangle, to the signal used by learner 2 to describe the same object?" The assumption was that signals, even if poorly learnt, were always associated with the meaning the were trying to express. Now that we have no meanings, how can we be sure which signal the learner was trying to replicate in their output?

Again, there is a measurement used in AGL studies which can help us: Global Similarity (Vokey & Brooks, 1992; Conway & Christiansen, 2005). Typically this measurement is used to give an idea of how similar training and test sets are to one another based on the number of fragments shared between signals in each set, *by first identifying the best possible alignment of those signals*[7]. To simplify the measure slightly and to keep it comparable to our previous analyses, instead of basing our calculation on the number of shared fragments between signals, we will base it on the nLD error values between signals. This is a purely cosmetic change, which ensures that the data, when graphed, shows error increasing or decreasing instead of similarity.

The way that this works is as follows: first we must find the number of elements by which a signal in generation *n* differs from its closest match in generation *n-1*. In other words, we calculate the nLD for all possible pairings of signals in the language, and then find the alignment between them that gives the lowest nLD score for each signal. This measure nicely captures our intuition that some particularly salient signals might be used more than once as the basis for generalisation in the recall round[8]. Once we have found this alignment, we can calculate the average nLD score for the whole language, as described in §3.4.3.

6.2.3 Results of Experiment V

---

[7] It is worth noting that for researchers designing AGL experiments, having training and test sets with high global similarity is undesirable. By insisting on selecting the alignment that maximises global similarity, they are therefore being highly conservative.

[8] This aspect does make the measure subtly different to that used in the previous studies, which assumed that each and every signal had to be used as the basis for generalisation for another signal once and once only. As a result, I also developed a slightly different metric to the standard global similarity measure described here, which matched this requirement by using a hill-climbing algorithm to identify the optimal alignment of signals (the alignment which gave us the lowest average nLD score, given the constraint that there must be a bi-unique mapping between signals). However, the results of applying this slightly more stringent bi-unique mapping version of global similarity to the experimental data was almost identical to the one obtained using the more standard metric, so is not included here.

Learnability Increases

The transmission error (measured as Global Similarity - see §6.3.2 for the description and exact modifications used) was calculated for each generation, and is shown in Figure 6.2 overleaf. If we examine all eight chains we find that although there is a lot of variation, there is an observable trend showing that error decreases over time. This can be seen more clearly in Figure 6.2.lower, which shows this same information, but in the form of a box-plot. Although there are a few outliers over the first few generations (one in the first generation who had notably poorer recall than the other participants in this condition, and another two learners in generations two and five who have notably better levels of recall), we can still see that the difference between the first and last learners appears substantial.

A one-way paired t-test was run on the first and final generations, which showed that error was significantly toward the end of the chains (mean decrease of 0.725; $t(7) = 4.6305$, $P < 0.002$) as compared to the beginnings. This result allows us to confirm the learnability hypothesis: languages are indeed becoming easier to learn over time. But how does this occur? It is to this question that we now turn.

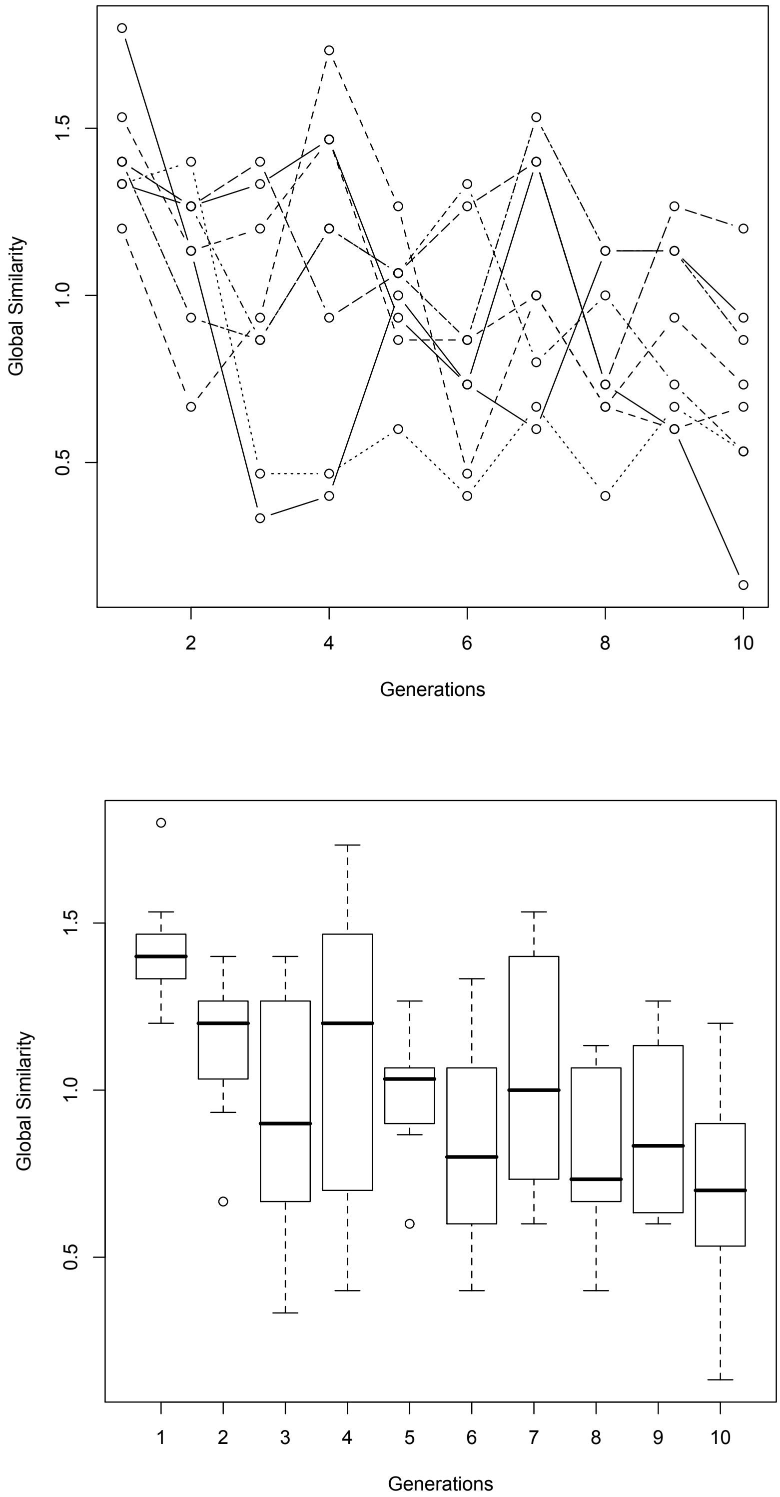


Figure 6.2: Global Similarity (error) decreases over generations in Experiment V. The upper graph shows every datapoint, whereas the lower graph shows the spread and descriptive statistics of the data

in more detail. Thick horizontal bars indicate the median, boxes show the inter-quartile ranges, and whiskers indicate the maximum and minimum values or 2 standard deviations away from the mean, whichever is smaller. Points outside this range are considered to be outliers. Here we find we have three points like this, in generations one, two and five.

Associated Chunk Strength Increases

In order to understand whether languages are becoming more structured over time, we first look to the ACS values. These provide us with an idea of how much re-use of fragments there is over time. These values are shown in Figure 6.3. From this we can see that there is a steady increase over time. If we run a one-way paired t-test on the values found at the beginning and the ends of each chain we find that this increase is massively significant (mean increase in ACS of 1.412; $t(7) = 6.203$, $P < 0.0003$). This tells us that the number of distinct fragments that appear and are repeated by learners has significantly increased over time. In the first generation, an average of just over 1 fragment per signal is repeated between generations; by the final generation, this has risen to nearly 3 fragments per signal. Given the fact that the average character length of signals is just 4.175 (a slight increase from the average of 4 in the initial input), this increase is substantial.

This indicates that at the very least, participants are combining individual characters in order to create chunks or fragments which seem to act like reusable units. Therefore, in some sense we can interpret the increase in ACS as an increase in structure as we defined it earlier. However, if we want to understand what this means in reality, we need to see what these languages actually look like.

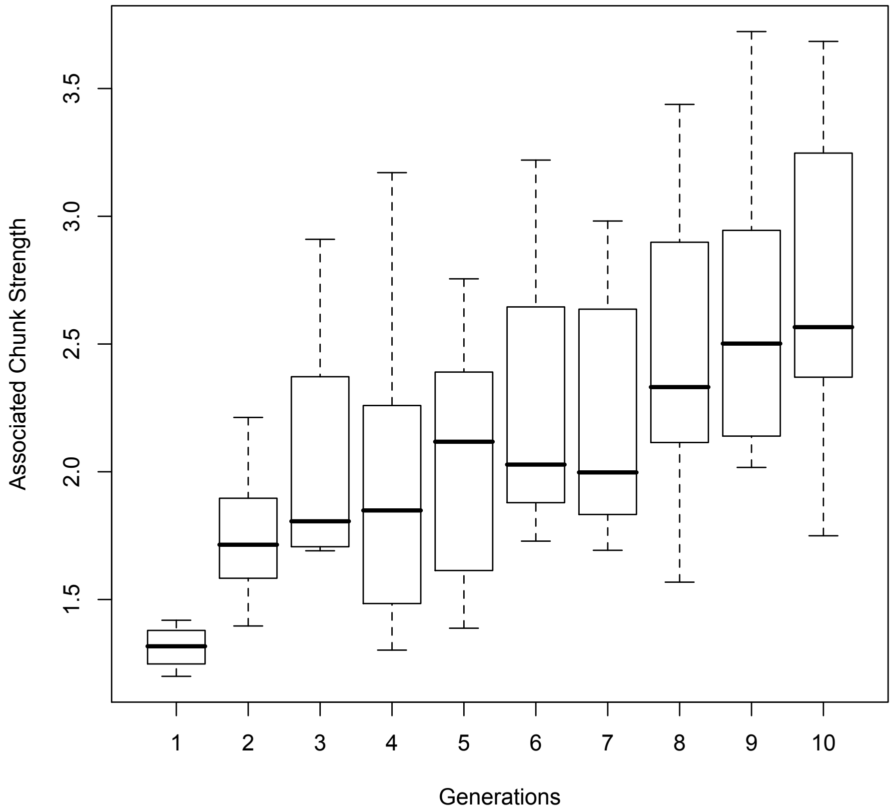


Figure 6.3: Associated Chunk Strength increases over generations in Experiment V. This tells us how many times on average the chunks (bigrams and trigrams) that form signals in the language appeared in the training data. The fact that this number increases over time indicates that these chunks are being treated as independent units by learners, and being reused in greater numbers as the experiment goes on.

Final Language Structures

Just as we did in Experiment I-IV, we can analyse the languages that emerge over time qualitatively by examining the signals we find in the final generation. Table 6.3 shows what the language depicted in Table 6.1 transformed into after being acquired by ten learners in succession. As you can see, there are a number of patterns that stand out in this system. Generally we can categorise strings as falling into distinct pattern types. The most striking of these are the four signals consisting of repeating triplets of characters. Next we find a group of four that all conform to the general pattern of **fcXaX**[9]. Some of these members overlap with another group which have the general pattern of **fcXXf**, so we can perhaps more parsimoniously describe these strings as being a type all together, starting with **fcXXX.** Finally, there are three strings that all contain the bigram **db**[10], and two short strings which appear to be unrelated to any of the others.

final language B
ccc
bbb
ddd
aaa
fcbab
fcbad
fcbaf
fcdaf
fcdcf
fcbcf
fdbdf
dbdaf
dbfbd
fca
cbd

*Table 6.3: An example of one of the more interesting languages found in the final generation of Experiment V. This language has a great deal of structure in it. We can identify four broad categories of strings emerging, each with minor variations: strings with three repeating characters; strings beginning with the bigram **fc;** strings containing the bigram **db**; and two irregulars. Crossing these boundaries is an additional pattern - that of palindromes. This feature was highly salient to learners, with four learners reporting it in their debriefing questionnaires.*

Another feature that is striking about this particular language is that many of the strings (53%) are palindromic (i.e. have the same structure forwards as they do backwards). Furthermore, this pattern seems to apply more globally across the language as a whole, rather than being a feature of any of the individual pattern types we have already identified. If we examine the raw data for this chain (Chain 2,

[9] The X's indicate where variant characters can appear.

[10] These share some similarities with some of the strings beginning **fcXXX**, but as the **db** combination does not appear anywhere else, it seems more parsimonious to analyse them as separate 'types'.

Appendix B5) we can see that this pattern emerges slowly, from no instances of palindromes in the initial language, to 2 in the first generation, 5 in generation five, 7 in generation seven, etc., until we reach 8 in generation 10. Although this chain was not the most learnable, four of the participants were consciously aware that this pattern existed[11]. This was the only example of this type of 'global pattern' to arise in any chain however, and the majority of participants only reported spotting local similarities between strings.

A more representative example of the kinds of languages that emerged is shown in Table 6.4. This shows the language at both the initial and the final generations. Here you can see that strings again seem to form groups that pattern together, but this time do not seem to have any higher organisational patterns. There does appear to be some internal structure to the signals however: certain chunks only appear at the end of strings (e.g. **-dad, -da, -ae**), whilst others only appear initially (e.g. **fec-, fce-, fed-**). The fact that these latter languages were learned much more easily than the initial languages appears to indicate that participants are sensitive to these patterns, and that they facilitate acquisition.

[11] Only one out of the four learners who noticed this pattern actually explained the system as working by having palindromes however (gen9). The remaining three reported this language as "having a rythym *[sic.]*" (gen7), "there is a pattern when you type - dum dum DUM dum dum" (gen10) and "you get lulled into the pattern on the keyboard...you need to leave your fingers on the keys so you can type quickly" (gen5).

initial language C	*final language C*
abecd	**bac**
acbf	**bfa**
aec	**dde**
bacdc	**eed**
bce	**fcada**
bfea	**fcdad**
cab	**fceae**
cbdae	**fceda**
cead	**fcede**
dbd	**fecad**
ddde	**fecae**
ebafb	**fecda**
eff	**fecdad**
fcfdf	**fedad**
fefa	**fedae**

Table 6.4: An example of one of the more typical languages found in the final generation of Experiment V. The distribution of fragments in the final language of chain C show many strong patterns. With the exception of the four three letter strings, there seem to be strong restrictions on which bigrams and trigrams start and end signals. As this final language was learned much more easily than the initial language, we can conclude that these patterns are aiding learners in their task.

6.2.4 Summary

This experiment was designed to try to isolate the effect of sequence memory constraints on cultural transmission. By modifying our previous experimental framework to remove as many biases as possible (including, most notably, the pressure being exerted upon signals to adapt to reflect the structure of the meanings) we attempted to answer the question of whether signals would still become more learnable and structured over time, even in the absence of any language-like task. To that end, initial languages consisting of carefully constructed letter-strings that contained as few regularities as possible were passed along a series of eight transmission chains.

Transmission error was calculated between each of the ten generations and found to decrease over time, indicating that the languages were becoming easier to learn. The change in distribution of bigrams and trigrams over time was investigated by measuring the associated chunk strength of fragments between generations. This revealed that certain chunks began to form and be re-used more often than others. Analysing the emergent languages descriptively revealed the appearance of patterns, particularly at the beginnings and ends of signals, that participants appeared sensitive to. Given the design of the experiment, the only explanation for the appearance of these distributional patterns is that they arise from sequential learning biases of the learners being amplified by cultural transmission.

6.3 Discussion of Experiment V

Although participants in Experiments I-IV are all given a learning, rather than a communication task, in Experiment V we have almost removed every trace of language. What, therefore, can this experiment tell us about how linguistic communication evolved? As the brief survey of the literature on sequence learning indicated (§6.1), several proponents have suggested that language has adapted to be learnable by domain-general, rather than language-specific, learning biases. In order to make this argument convincing, we require evidence not only that the domain-general learning biases that humans possess can more easily *process* certain types of structure, but that those same domain-general learning biases can actually *generate* such structures. This experiment provides that evidence.

The first thing to note, therefore, is that in spite of the fact that little if anything is linguistic about this study, we find similar outcomes arising here as we do with the other more obviously linguistic experiments. This is interesting, as it indicates that the same underlying process is at work in both cases. We find cumulative adaptation resulting in better recall over time. The second point that can be made concerns the point at which we started the previous experiments. In constructing our initial languages for Experiments I-IV, we sampled from a set of pre-generated CV-syllables. The reason we did that was because we know that language is, in a sense, composed of these low-level units which have not been assigned with a meaning[12]. In order to create meaningful distinctions at higher levels of organisation, we require these 'worker units' to do the hard graft. But where do they come from? How do they get their structure? This experiment can perhaps speak to questions like these.

To recap, when we transmitted signals without any meanings, what we got out at the end was the appearance of small chunks that began to get reused with greater frequency amongst all the signals in the language. These chunks were seen to emerge because of domain-general constraints on sequence memory being amplified by the process of cultural transmission. Language is full of meaningless units such as these. An implication from this work then is that these constraints could have been an important factor which shaped linguistic structure. If this is the case, there is no reason to suspect that this process was limited only to a stage in language evolution where there were only meaningless signals. Sequence learning biases are likely still operating and exerting a subtle force on language today, but as with the case of the creative capacities of children only being evident in cases of new language formation, and not when acquiring a fully-fledged language (Senghas & Coppola, 2001; see §3.2.2), we may only witness these effects in unusual circumstances.

[12] Note that I am not suggesting that a syllable is inherently a meaningless unit - of course, a syllable can sometimes be a whole word, or even convey an entire proposition. Duh.

Chapter Seven

Conclusions

7.1 Looking Back: In Answer to Earlier Questions

We began this thesis with three questions in mind. Namely:

1. Why is language structured the way it is and not some other way?
2. How does the process of cultural transmission give rise to language structure?
3. Can features of language structure which appear to be designed for communication evolve in the absence of a) actual communication, and b) intentional design?

We should now be in a position to provide some answers. This section will tackle each of these questions in turn.

7.1.1 Why is language structured the way it is and not some other way?

In Chapters 2 and 3 I reviewed literature which suggested a partial answer to the first question: language has the structural features it has because those are the features that emerge when initially unstructured systems get culturally transmitted via iterated learning. In other words, since language is a complex adaptive system in its own right, capable of evolving culturally, these processes of cultural transmission could have an explanatory role to play in understanding the emergence of structure.

Although this intuition had previously been verified as being logically sound in computer simulations and mathematical models, the experiments described in Chapters 4-6 demonstrate that the concept is applicable to human learners. These studies show that language-like structural relationships can emerge in initially unstructured artificial languages, when they are culturally acquired via iterated learning along linear chains of human participants. Importantly, we can demonstrate a direct link between specific constraints being placed on languages during transmission and the structures that emerge. In other words, we can account for why the artificial languages were structured the way they were and not some other way, by specifically identifying and manipulating these pressures to different effects.

In particular, Experiments I-IV show that the signals in a language always adapt to reflect the structure of the meanings that they express. However, the relationship between signals and meanings is not simple. In Experiments I and II when there were only pressures acting on systems to be learnable, we found that the meaning-space became reorganised in response to this requirement. Meanings became underspecified by signals. Importantly, however, this underspecification was adaptive: rather than affecting meanings and signals at random, there was a kind of systematic levelling or recategorisation process at work which led to semantically related objects being given the same signal. This made these systems very easy to learn and reproduce in full, even when participants were only being trained on half of the data (Chapter 4). In Experiments III and IV on the other hand, when there were pressures being placed on the system to not only be learnable, but also expressive, a different kind of structure emerged. Here signals evolved to express meanings, or parts of meanings, compositionally.

In Experiment V there were no meanings for signals to adapt to. In this study (described in Chapter 6) many elements of the previous studies were stripped out, leaving a task that was not remotely language-like. The purpose was to investigate whether sequence memory biases could give rise to structure in signals. This was indeed found to be the case. From an initial starting point of string-sets that contained very little regularity or repetition of sequences of characters, the systems evolved to have increasingly learnable distributional structures.

In sum, the languages in our experiments came to have the structures that they did because they were adapting to pressures arising from transmission. Some of these pressures were external to the learners and imposed upon the language without their knowledge (e.g. the semantic bottleneck, filtering) and others were internal to the learners (e.g. biases on sequence learning). As Chapter 4 discussed, these pressures get exerted at different points in the transmission cycle, and only some of them are directly controllable by the experimenter. Therefore, it is important to remember that although we have focused our explanations on the three kinds of bottleneck named above, there are others at work also. For instance, we would predict that constraints on sequence memory would also be playing a role in shaping the languages in Experiments I and IV, even though we cannot directly detect it.

7.1.2 How does the process of cultural transmission give rise to language structure?

Turning to the question of exactly *how* processes of cultural transmission give rise to language structure, we can use the fact that we have the entire recorded histories of all of the languages in the chains to track the evolution of individual signal forms, or in the case of compositional languages, parts of signals, over time. This kind of analysis technique was demonstrated in Chapters 4 and 5, and revealed that there were clear lines of descent between signals. The amount of signal variants was found to decrease over time as a result of competition. In this way, the system slowly emerged.

One of the key findings of the mathematical models of iterated learning described in Chapters 3 and 5 suggests that cultural transmission works to amplify the prior learning biases of agents, thus giving rise to structure. We saw in Experiment III how constraints on cultural transmission (in this particular case, a combination of the semantic bottleneck and filtering) increase the amount of regularity in the language as a whole, in virtue of the participants' training data containing more structure locally. Generalisation based upon this locally more regular structure leads to increasing structure globally.

These same mathematical models also make different predictions about the role of cultural transmission in processes of iterated learning. In one model (Griffiths & Kalish, 2005; 2007) the outcome of iterated learning has been shown to be just a reflection of the learning biases of the agents. In other words, cultural transmission is not adding anything to the process or leaving any mark on the resulting system that was not already, in a sense, present in the mind of the agents *a priori*. Other models show that by altering the way agents select between competing hypotheses about the data (Kirby *et al.,* 2007), or by changing the population dynamics of the models (Ferdinand & Zuidema, 2009), the outcome of iterated learning *is* modified by cultural transmission. The experiments support this latter idea that transmission is adding something. As an example, the fact that we found a difference between the results of our first experiment (which had no requirement to be expressive) and our third experiment (which did have a requirement to be expressive) shows that it is the manipulation of the way languages were being transmitted which is responsible for the effect, and not just the learning biases of participants alone.

7.1.3 Can features of language structure which appear to be designed for communication evolve in the absence of a) actual communication, and b) intentional design?

None of the experiments contain any communicative element to them. Participants were not learning the languages in an interactive environment or using the languages 'for' anything. In Experiment V in particular, the stimuli was not even referred to as a language; participants were recruited on the understanding that they were to take part in a recall experiment. Yet nevertheless all experiments show that the transmitted systems adapt over time and become structured. As participants are not actually using the language for communication, how can we be sure that the structure that appears is in fact of the type that is useful for communication? At least in Experiments I-IV, it is important to note that the types of structure that arise (underspecification and compositionality) are widely found in human language. Here we know that they underlie communication - compositionality, in allowing for greater productivity in language, and underspecification in allowing objects to be categorised together and assigned a common label (for example, the common noun 'chair'). The structures that arose in Experiment V are harder to interpret, as it is difficult to know what the letters in the signal strings correspond to in language. Nevertheless, the fact that learnability improves over time is an indicator that the signals at the end of the study would be better candidates to be used as labels if meanings were suddenly introduced than the signals at the beginning of the study.

In terms of the signals evolving in the absence of intentional design, again we can look to the contrast between the results found in Experiment I and Experiment III. The filtering condition was invisible to participants. Even if they had been making intentional changes to the language in the first experiment (for instance, by choosing to only try to memorise difficult signals, or ignore minor variations, or use mnemonic tricks for recall, or by having any goal other than straightforward reproduction of the signal-meaning pairs), they would have had no way of knowing to perform a different action in the third experiment. Therefore, we can safely conclude that the differences we saw between conditions were not the result of any goal directed behaviour by learners.

7.2 Implications for Language Evolution

As the review in Chapter 2 hopefully highlighted, the field of language evolution has a number of divisions within its ranks. This is a good sign of healthy debate. Broadly speaking, theories of language evolution can be separated along two main lines: on the issue of language-specific learning biases, and on the role of cultural transmission. On the one hand there are researchers who hold that language must be reliant on innate knowledge specific to language (e.g. Chomsky, 1965; Pinker, 1994), whereas others stress the importance of general cognitive mechanisms (e.g. Elman *et al.*, 1996; Christiansen & Chater, 2008). Separately, there are also researchers divided over the importance of cultural transmission in explanations of language evolution, with some placing no emphasis on it (e.g. Pinker & Bloom, 1990) and others who contend that it actually does some work (e.g. Brighton *et al.*, 2005). What, if anything, do these experiments contribute to our understanding of these issues?

The results of these studies on human learners do not, on their own, tell us anything about the nature of learning biases involved in language. It could be argued that the fact that we see language-like structures emerging here is simply a reflection of the underlying linguistic capabilities that is the biological legacy of *Homo sapiens* everywhere. If this is the conclusion that some readers draw from this work, then it is one I can *just about* live with. However, what cannot be in doubt is the fact that constraints on cultural transmission are actively 'doing something' here. The participants in different experimental conditions did not have different processing mechanisms: what shaped the different structural outcomes was the data that was being transmitted and how it was affected by the external manipulations we made. Therefore, one thing that these studies confirm without a doubt is that theories of language evolution need to take cultural evolution more seriously.

I said that I could live with the reader coming away with the conclusion that humans have language-specific learning biases. That is not to say that I think that is the right conclusion however. The point cannot be established based on these results alone, but needs to be understood in the wider empirical context. The starting point for all of these studies were findings coming from computer simulations of iterated learning. In these models, agents are not rewarded for successful communication. They have no pre-existing language. They have no language-specific learning biases. What they do have are general cognitive mechanisms that allow them to process sequences, the ability to form categories and make generalisations, and the willingness to copy others.

Given that the experiments with human learners essentially replicate the behaviour of these much simpler agents, it seems that the most parsimonious explanation is that universal structural properties of language do not require language-specific brain mechanisms. If we follow this argument to its natural conclusion then, one implication that we can take away from these studies is that compositional language did not necessarily require much to get off the ground. As long as there is a basic desire to distinguish between different objects (a need for expressivity), and some desire to copy the vocalisations of others (a need for learnability), cultural evolution will deliver.

7.3 Key Contributions

The key findings of the five studies tell us a number of interesting things. Firstly, that it is possible to witness the cultural evolution of language in the laboratory; secondly, that results from computer simulations of the process can, to a large extent, be said to generalise to human learners; and thirdly, that language adapts to those pressures placed upon it during transmission. The importance of these three results should not be underestimated.

Evolutionary linguistics is a field that has traditionally suffered (or at least been perceived to suffer) from a lack of data. These experiments provide a new way for us to extract information relevant to understanding the processes that underlie the emergence of language-like systems. It is hoped that the development of this experimental methodology will open the door for more research in this area. Indeed, the early signs give us reason for optimism. Several studies which acknowledge this framework have already been conducted, extending the work presented here by: exploring different population structures (Winters, 2009; Line, 2010); manipulating characteristics of the meaning-spaces (Beqa *et al.*, 2008; Matthews *et al*, 2010);

investigating regularisation (Smith & Wonnacott, 2010); exploring different modalities (Tamariz, Brown & Murray 2010); and even attempting to compare the performance of adults and children (Flaherty & Kirby, 2008). The recent growth in this area has recently been charted in a *Trends in Cognitive Science* paper (Scott-Phillips & Kirby, 2010).

That the results of these experiments also support computational simulations of the process enables us to not only respond to critics of the modelling approach and make our findings more accessible to researchers in other fields, but it also enables us to better understand the nature of the cognitive mechanisms responsible for the appearance of structure in these systems. Likewise, where results deviate from those predicted by simulations, it serves to highlight areas where our modelling assumptions are incorrect. Thus I hope to have demonstrated that both research methodologies are mutually supportive, and have a greater impact when their results are viewed together rather than individually.

Finally, the discovery that language does in fact adapt to constraints arising during transmission adds something concrete to our understanding of how language might have evolved in our species; namely, that biological evolution is not the only adaptive mechanism capable of generating linguistic structure. This should not be taken as suggesting that biological adaptations play an insignificant role in language emergence, however. One of the themes that recurs throughout the thesis is that cultural transmission works to amplify any biases present either in the learner, or arising from the transmission process itself. In some sense, the outcome of this work is the generation of more questions. Where do learning biases come from? And also: what neurological mechanisms support iterated learning in humans? How did those mechanisms evolve? These are unfortunately issues that must be left for future investigation.

References

Aiello, L. C. & Dunbar, R. I. M. (1993), 'Neocortex size, group size and the evolution of language', *Current Anthropology* **34**, 184–93.

Allport, G. W. & Postman, L. J. (1947), *The Psychology of Rumour*, New York: Henry Holt.

Anderson, H. (1973), 'Abductive and deductive change', *Language* **40**, 765–793.

Arensburg, B., Tillier, A. M., Vandermeersch, B., Duday, H., Schepartz, L. A. & Rak, Y. (1989), 'A middle paleolithic human hyoid bone', *Nature* **338**, 758–760.

Atran, S. (2001), 'The trouble with memes', *Human Nature* **12**, 351–381.

Aunger, R. (2002), *The Electric Meme: a new theory of how we think*, New York: The Free Press.

Aunger, R., ed. (2000), *Darwinizing Culture: The Status of Memetics as a Science.*, Oxford: Oxford University Press.

Bangerter, A. (2000), 'Transformation between scientific and social representations of conception', *British Journal of Social Psychology* **39**, 521–535.

Barrett, L., Dunbar, R. & Lycett, J. (2002), *Human Evolutionary Psychology*, London: Palgrave.

Bartlett, F. (1932), *Remembering: a study in experimental and social psychology*, Cambridge: Cambridge University Press.

Barton, N. H. (2007), *Evolution*, Cold Spring Harbour, NY: Cold Spring Harbour Press.

Basalla, G. (1988), *The evolution of technology*, Cambridge: Cambridge University Press.

Batali, J. (1998), Computational simulations of the emergence of grammar, *in* J. Hurford, M. Studdert-Kennedy & C. Knight, eds, 'Approaches to the evolution of language: Social and cognitive bases', Cambridge: Cambridge University Press, pp. 405–426.

Batali, J. (2002), The negotiation and acquisition of recursive grammars as a result of competition among exemplars, *in* T. Briscoe, ed., 'Linguistic evolution through language acquisition: Formal and computational models', Cambridge: Cambridge University Press, pp. 111–172.

Bates, E., Thal, D., Finlay, B. & Clancy, B. (2003), Early language development and its neural correlates, *in* I. Rapin & S. Segalowitz, eds, 'Child Neurology – Volume 6 of the Handbook of Neuropsychology.', 2nd edn, Amsterdam: Elsevier.

Baum, W. M., Richerson, P., Efferson, C. M. & Paciotti, B. M. (2004), 'Cultural evolution in laboratory microsocieties including traditions of rule giving and rule following', *Evolutionary Human Behavior* **25**, 305–326.

Bechtel, W. & Abrahamsen, A. (2002), *Connectionism and the Mind: Parallel Processing, Dynamics and Evolution in Networks*, 2nd edn, Oxford: Blackwell.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N., Holland, J., Ke, J., Larsen-Freeman, D. & Schoenemann, T. (2009), 'Language is a complex adaptive system: Position paper', *Language Learning* **51**(S1), 1–26.

Bengtson, J. D. & Ruhlen, M. (1994), Global etymologies, *in* M. Ruhlen, ed., 'Pn the origin of languages: studies in linguistic taxonomy', Stanford: Stanford University Press, pp. 277–336.

Benton, T. (2000), Social causes and natural relations, *in* H. Rose & S. Rose, eds, 'Alas, Poor Darwin', New York: Harmony, pp. 206–224.

Beqa, A., Kirby, S. & Hurford, J. (2008), Regular morphology as a cultural adaptation: non-unifomr frequency in an experimental iterated learning model, *in* A. Smith, K. Smith & R. F. i Cancho, eds, 'The Evolution of Language: Proceedings of the 7th International Conference (EVOLANG7)', Singapore: World Scientific, pp. 401–402.

Bickerton, D. (1981), *Roots of Language*, Ann Arbor, MI: Karoma.

Bickerton, D. (2003), Symbol and structure: a comprehensive framework for language evolution, *in* M. H. Christiansen & S. Kirby, eds, 'Language Evolution: The States of the Art', Oxford: Oxford University Press, pp. 77–93.

Bickerton, D. (2007), 'Language evolution: A brief guide for linguists', *Lingua* **117**, 510–526.

Biro, D., Inoue-Nakamura, N., Tonooka, R., Yamakoshi, G., Sousa, C. & Matsuzawa, T. (2003), 'Cultural innovation and transmission of tool use in wild chimpanzees: evidence from field experiments', *Animal Cognition* **6**, 213–223.

Blackmore, S. (1999), *The Meme Machine*, Oxford: Oxford University Press.

Bloch, M. (2000), A well-disposed social anthropologist's problems with memes, *in* R. Aunger, ed., 'Darwinizing Culture: The Status of Memetics as a Science', Oxford: Oxford University Press, pp. 189–204.

Boe, L.-J., Heim, J.-L., Honda, K. & Maeda, S. (2002), 'The potential neanderthal vowel space was as large as that of modern humans', *Journal of Phonetics* **30**, 465–484.

Boesch, C. (1996), 'The emergence of cultures among wild chimpanzees', *Proceedings of the British Academy* **88**, 251–268.

Boesch, C. & Tomasello, M. (1998), 'Chimpanzee and human cultures', *Current Anthropology* **39**(5), 591–614.

Boyd, R. & Richerson, P. (1985), *Culture and the evolutionary process*, Chicago: University of Chicago Press.

Boyd, R. & Richerson, P. (1995), 'Why does culture increase human adaptability?', *Ethology and Sociobiology* **16**(2), 125–143.

Brighton, H. (2002), 'Compositional syntax from cultural transmission', *Artificial Life* **8**(1), 25–54.

Brighton, H. (2003), Simplicity as a driving force in linguistic evolution, PhD thesis, The University of Edinburgh.

Brighton, H., Smith, K. & Kirby, S. (2005), 'Language as an evolutionary system', *Physics of Life Reviews* **2**, 177–226.

Brown, R. (1973), *A First Language: The Early Stages*, Cambridge, MA: Harvard University Press.

Brown, R. & Hanlon, C. (1970), Derivational complexity and order of acquisition in child speech, *in* J. R. Hayes, ed., 'Cognition and the development of language', New York: Wiley.

Bryant, J. M. (2004), 'An evolutionary social science? a skeptic's brief, theoretical and substantive', *Philosophy of the Social Sciences* **34**, 451–492.

Bullock, S. & Noble, J. (2000), 'Evolutionary simulation modelling clarifies interactions between parallel adaptive processes', *Behavioral and Brain Science* **23**(1), 150–151.

Caldwell, C. A. & Millen, A. E. (2008), 'Experimental models for testing hypotheses about cumulative cultural evolution', *Evolutionary Human Behavior* **29**, 165–171.

Caldwell, C. A. & Millen, A. E. (2010), 'Conservatism in laboratory microsocieties: unpredictable payoffs accentuate group-specific traditions', *Evolutionary Human Behavior* **31**, 123–130.

Caldwell, C. A. & Whiten, A. (2006), Social learning in monkeys and apes: Cultural animals?, *in* C. J. Campbell, K. C. Fuentes, K. C. MacKinnon, M. Panger & S. Bearder, eds, 'Primates in Perspective', Oxford: Oxford University Press.

Caldwell, C. & Whiten, A. (2004), 'Testing for social learning and imitation in common marmosets, callithrix jacchus, using as artificial fruit', *Animal Cognition* **7**, 77–85.

Cangelosi, A. & Parisi, D. (1998), 'The emergence of a 'language' in an evolving population of neural networks', *Connection Science* **10**(2), 83–97.

Cavalli-Sforza, L. L. (2000), *Genes, People, and Languages*, Berkeley and Los Angeles: University of California Press.

Cavalli-Sforza, L. L. & Feldman, M. (1981), *Cultural transmission and Evolution:a quantitative approach*, Princeton University Press.

Cheney, D. L. & Seyfarth, R. M. (1990), *How monkeys see the world: inside the mind of another species*, Chicago: University of Chicago Press.

Chomsky, N. (1959), 'A review of skinner's "verbal behavior"', *Language* **35**, 26–58.

Chomsky, N. (1965), *Aspects of the theory of syntax*, Cambridge, MA: MIT Press.

Chomsky, N. (1975), *Reflections on Language*, New York: Pantheon Books.

Chomsky, N. (1980), *Rules and Representations*, Oxford: Basil Blackwell.

Chomsky, N. (1986), *Knowledge of Language*, New York: Praeger.

Chomsky, N. (1988), *Language and Problems of Knowledge*, Cambridge, MA: MIT Press.

Chomsky, N. (1991), Linguistics and cognitive science: Problems and mysteries, *in* A. Kasher, ed., 'In the Chomskyan Turn', Oxford: Basil Blackwell, pp. 26–53.

Chomsky, N. (2002), *On Nature and Language*, Cambridge: Cambridge University Press.

Christiansen, M. H. (2000), Using artificial language learning to study language evolution: exploring the emergence of word-order universals, *in* J. Desalles & L. Ghadakpour, eds, 'The Evolution of Language 3rd International Conference', pp. 45–48.

Christiansen, M. H. & Chater, N. (2008), 'Language as shaped by the brain', *Behavioral and Brain Science* **31**(5), 489–509.

Christiansen, M. H. & Devlin, J. T. (1997), Recursive inconsistencies are hard to learn: A connectionist perspective on universal word order correlations, *in* 'Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society', Mahwah, NJ: Lawrence Erlbaum Associates, pp. 113–118.

Christiansen, M. H. & Kirby, S. (2003), Language evolution: The hardest problem in science?, *in* 'Language Evolution: The States of the Art', Oxford: Oxford University Press.

Clark, A. (1998), Magic words: How language augments human computation, *in* P. Carruthers & J. Boucher, eds, 'Language and Thought: Interdisciplinary Themes', Cambridge: Cambridge University Press, pp. 162–183.

Clark, A. (2006), 'Language, embodiment, and the cognitive niche', *Trends in Cognitive Science* **10**(8), 370–374.

Clark, H. (1996), *Using Language*, Cambridge: Cambridge University Press.

Cooper, R. (2002), *Modelling High-Level Cognitive Processes*, Mahwah, NJ: Lawrence Erlbaum Associates.

Cornish, H. (2005), The role of meaning within the iterated learning model of language evolution. Honours Dissertation, The University of Edinburgh.

Cornish, H. (2006), Iterated learning with human subjects: an empirical framework for the emergence and cultural transmission of language, Master's thesis, The University of Edinburgh.

Cornish, H. (2010), 'Investigating how cultural transmission leads to the appearance of design without a designer in human communication systems', *Interaction Studies* **11**(1), 112–137.

Cornish, H., Tamariz, M. & Kirby, S. (2009), 'Complex adaptive systems and the origins of adaptive structure: What experiments can tell us', *Language Learning* **59**(s1), 187–205.

Cowie, F. (1999), *What's Within: Nativism Reconsidered*, Oxford: Oxford University Press.

Crain, S. & Pietroski, P. (2001), 'Nature, nurture and universal grammar', *Linguistics and Philosophy* **24**, 139–186.

Croft, W. (2000), *Explaining language change: An evolutionary approach*, London: Longman.

Curtiss, S. (1977), *Genie: A psycholinguistic study of a modern-day 'wild child'*, New York: Academic Press.

Custance, D., Whiten, A., Sambrook, T. & Galdikas, B. (2001), 'Testing for social learning in the 'artificial fruit' processing of wildborn orangutans (pongo pygmaeus), tanjung putin, indonesia', *Animal Cognition* **4**, 305–313.

Darwin, C. (1859), *On the Origins of Species*, London: John Murray.

Darwin, C. (1871/2004), *The Descent of Man*, London: Penguin Classics.

Dawkins, R. (1976), *The Selfish Gene*, 30th anniversary edn, Oxford: Oxford University Press.

de Beule, J. & Bergen, B. K. (2006), On the emergence of compositionality, *in* A. Cangelosi, A. Smith & K. Smith, eds, 'Proceedings of the 6th International Conference on the Evolution of Language', pp. 35–42.

de Boer, B. (2005), 'Evolution of speech and its acquisition', *Adaptive Behavior* **13**, 281–292.

Deacon, T. W. (1997), *The symbolic species: The co-evolution of language and the brain*, New York: W.W. Norton.

Dediu, D. & Ladd, D. R. (2007), 'Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, aspm and microcephalin', *Proceedings of the National Academy of Sciences of the United States of America* **104**(26), 10944–10949.

Dennett, D. (1995), *Darwin's Dangerous Idea: evolution and the meanings of life*, London: Penguin Press.

Dennett, D. & McKay, R. (2006), 'A continuum of mindfulness', *Behavioral and Brain Science* **29**, 353–354.

Dindo, M., Whiten, A. & de Waal, F. B. M. (2009), 'In-group conformity sustains different foraging traditions in capuchin monkeys (cebus apella)', *Public Library of Science One* **4**(11), e7858.

DiPaolo, E. (1997), 'An investigation into the evolution of communication', *Adaptive Behavior* **6**, 285–324.

Doupe, A. & Kuhl, P. (1999), 'Birdsong and human speech: Common themse and mechanisms', *Annual Review of Neuroscience* **22**, 567–631.

Dunbar, R. (1996), *Grooming, Gossip and the Evolution of Language*, Cambridge, MA: Harvard University Press.

Durham, W. H. (1991), *Coevolution: Genes, Culture, and Human Diversity*, Stanford: Stanford University Press.

Ellefson, M. & Christiansen, M. H. (2000), Subjacency constraints without universal grammar: evidence from artificial language learning and connectionist modeling, *in* L. Gleitman & A. Joshi, eds, 'Proceedings of the 22nd Annual Conference of the Cognitive Science Society', pp. 645–650.

Ellison, M. & Kirby, S. (2006), Measuring language diversion by intra-lexical comparison, *in* 'Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL', Association for Computational Linguistics, pp. 273–280.

Elman, J. L. (1990), 'Finding structure in time', *Cognitive Science* **14**(2), 179–211.

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

Elman, J. L., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996), *Rethinking Innateness: A connectionist perspective on development*, Cambridge, MA: MIT Press.

Evans, N. & Levinson, S. C. (2009), 'The myth of language universals: Language diversity and its importance for cognitive science', *Behavioral and Brain Science* **32**, 429–448.

Everett, D. L. (2005), 'Cultural constraints on grammar and cognition in piraha', *Current Anthropology* **46**(4), 621–646.

Falk, D. (1980), 'A reanalysis of the south african australopithecine natural endocasts', *American Journal of Physical Anthropology* **43**, 123–132.

Feher, O., Wang, H., Saar, S., Mitra, P. & Tchernichovski, O. (2009), 'De novo establishment of wild-type song culture in the zebra finch', *Nature* **459**, 564–568.

Ferdinand, V. & Zuidema, W. (2009), Thomas' theorem meets bayes' rule: a model of the iterated learning of language, *in* 'Proceedings of the 31st Annual Conference of the Cognitive Science Society', pp. 2974–2979.

Fitch, W. T. (2000), 'The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals', *Phonetica* **57**, 205–218.

Fitch, W. T. (2010), *The Evolution of Language*, Cambridge: Cambridge University Press.

Fitch, W. T. & Hauser, M. (2004), 'Computational constraints on syntactic processing in a non-human primate', *Science* **303**, 377–380.

Fitch, W. T., Hauser, M. D. & Chomsky, N. (2005), 'The evolution of the language faculty: Clarifications and implications', *Cognition* **97**(2), 179–210.

Flaherty, M. & Kirby, S. (2008), Iterated language learning in children, *in* A. Smith, K. Smith & R. F. i Cancho, eds, 'The Evolution of Language: Proceedings of the 7th International Conference (EVOLANG7)', pp. 425–426.

Flynn, E. (2008), 'Investigating children as cultural magnets: do young children transmit redundant information along diffusion chains?', *Philosophical Transactions of the Royal Society B* **363**, 3541–3551.

Flynn, E. & Whiten, A. (2008), 'Cultural transmission of tool-use in young children: a diffusion chain study', *Social Development* **17**, 699–718.

Fodor, J. A. (1983), *Modularity of Mind*, Cambridge, MA: MIT Press.

Fox, A. (1995), *Linguistic Reconstruction: An Introduction to Theory and Method*, Oxford: Oxford University Press.

Francis, N. & Kucera, H. (1982), *Frequency analysis of English usage: lexicon and grammar*, Boston, MA: Houghton Mifflin.

Galantucci, B. (2005), 'An experimental study of the emergence of human communication systems', *Cognitive Science* **29**, 737–767.

Galantucci, B. & Garrod, S. (2010), 'Experimental semiotics: A new approach for studying the emergence and the evolution of human communication', *Interaction Studies* **11**(1), 1–13.

Gardner, R. A. & Gardner, B. T. (1969), 'Teaching sign language to a chimpanzee', *Science* **165**, 644–672.

Garrod, S. & Anderson, A. (1987), 'Saying what you mean in dialogue: a study in conceptual and semantic co-ordination', *Cognition* **27**, 181–218.

Garrod, S. & Doherty, G. (1994), 'Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions', *Cognition* **58**, 181–215.

Garrod, S., Fay, N., Lee, J., Oberlander, J. & MacLeod, T. (2007), 'Foundations of representation: where might graphical symbol systems come from?', *Cognitive Science* **31**(6), 961–987.

Garrod, S., Fay, N., Walker, B. & Swoboda, N. (2010), 'Can iterative learning explain the emergence of graphical symbols?', *Interaction Studies* **11**, 33–50.

Gell-Mann, M. (1992), Complexity and complex adaptive systems, *in* J. Hawkins & M. Gell-Mann, eds, 'The Evolution of Human Languages', Addison-Wesley: Reading, MA.

Gell-Mann, M. (1994), *The Quark and the Jaguar: Adventures in the simple and the complex*, New York: W.H. Freeman and Co.

Gergely, G. & Csibra, G. (2006), Sylvia's recipe: The role of imitation and pedagogy in the transmission of human culture, *in* N. J. Enfield & S. C. Levinson, eds, 'Roots of Human Sociality: Culture, Cognition, and Human Interaction', Oxford: Berg Publishers, pp. 229–255.

Gerrard, R. W., Kluckhohn, C. & Rapoport, A. (1956), 'Biological and cultural evolution', *Behavioral Science* **1**, 6–34.

Gil, D. (2008), How much grammar does it take to sail a boat? (or, what can material artifacts tell us about the evolution of language?), *in* A. Smith, K. Smith & R. F. i Cancho, eds, 'Proceedings of the 7th International Conference on the Evolution of Language', Singapore: World Scientific, pp. 123–130.

Gold, E. (1967), 'Language identification in the limit', *Information and Control* **16**, 447–474.

Goldin-Meadows, S. & Mylander, C. (1998), 'Spontaneous sign systems created by deaf children in two cultures', *Nature* **391**, 279–281.

Gomez, R. L. & Gerken, L. A. (2000), 'Infant artificial language learning and language acquisition', *Trends in Cognitive Science* **4**(5), 178–186.

Gould, S. (1997), 'Darwinian fundamentalism', *New York Review of Books* **44**(10), 34–37.

Gould, S. & Vrba, E. (1982), 'Exaptation: a missing term in the science of form', *Paleobiology* **8**(1), 4–15.

Griffiths, T., Christian, B. R. & Kalish, M. (2008), 'Using category structures to test iterated kearning as a method for identifying inductive biases', *Cognitive Science* **32**, 68–107.

Griffiths, T. & Kalish, M. (2005), A bayesian view of language evolution by iterated learning, *in* 'Proceedings of the 27th Annual Conference of the Cognitive Science Society'.

Griffiths, T. & Kalish, M. (2007), 'Language evolution by iterated learning with bayesian agents', *Cognitive Science* **31**(3), 441–480.

Griffiths, T., Kalish, M. & Lewandowsky, S. (2008), 'Theoretical and experimental evidence for the impact of inductive biases on cultural evolution', *Philosophical Transactions of the Royal Society B* **363**, 3469–3476.

Guglielmino, C. R., Viganotti, C., Hewlett, B. & Cavalli-Sforza, L. L. (1995), 'Cultural variation in africa: Role of mechanisms of transmission and adaptation.', *Proceedings of the National Academy of Sciences of the United States of America* **87**.

Haiman, J. (1980), 'The iconicity of grammat: Isomorphism and motivation', *Language* **56**, 515–540.

Hallpike, C. R. (1986), *The principles of social evolution*, Oxford: Clarendon Press.

Hare, M. & Elman, J. L. (1995), 'Learning and morphological change', *Cognition* **56**, 61–98.

Hashimoto, T. (2002), The constructive apprach to the dynamical view of language, *in* A. Cangelosi & D. Parisi, eds, 'Simulating the evolution of language', Cambridge: Cambridge University PressC.

Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002), 'The faculty of language: What is it, who has it, and how did it evolve?', *Science* **298**, 1569–1579.

Hauser, M. D. & Fitch, T. (2003), What are the uniquely human components of the language faculty?, *in* M. H. Christiansen & S. Kirby, eds, 'Language Evolution: The States of the Art', Oxford: Oxford University Press.

Healey, P. G. T., Swoboda, N., Umata, I. & King, J. (2007), 'Graphical language games: Interactional consrtaints on representational form', *Cognitive Science* **31**(2), 285–309.

Hein, J., Schierup, M. H. & Wiuf, C. (2005), *Gene genealogies, variation and evolution: A primer in coalescent theory*, Oxford: Oxford University Press.

Hendriks, P. (2000), The problem with the logical problem of language acquisition, *in* L. Gleitman & A. Joshi, eds, 'Proceedings of the 22nd Annual Conference of the Cognitive Science Society', Mahwah, NJ: Lawrence Erlbaum Associates.

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

Hirsh-Pasek, K., Treiman, R. & Schneiderman, M. (1984), 'Brown and hanlon revisited: mother's sensitivity to ungrammatical forms', *Journal of Child Language* **11**, 81–88.

Hockett, C. F. (1958), *A course in modern linguistics*, New York: Macmillan.

Hockett, C. F. & Altmann, S. A. (1968), A note on design features, *in* T. A. Sebrook, ed., 'Animal Communications: Techniques of study and result of research', Bloomington: Indiana University Press, pp. 61–72.

Hoefler, S. (2006), Why has ambiguous syntax emerged?, *in* A. Cangelosi, A. Smith & K. Smith, eds, 'Proceedings of the 6th International Conference on the Evolution of Language', pp. 123–130.

Hoefler, S. (2009), Modelling the Role of Pragmatic Plasticity in the Evolution of Linguistic Communication, PhD thesis, The University of Edinburgh.

Holden, C. & Mace, R. (1999), 'Sexual dimorphism in stature and women's work: A phylogenetic cross-cultural analysis', *American Journal of Physical Anthropology* **10**, 27–45.

Holland, J. (1995), *Hidden Order: how adaptation builds complexity*, Reading, MA: Addison-Wesley.

Horner, V., Proctor, D., Bonnie, K., Wh, A. & de Waal, F. B. M. (2010), 'Prestige affects cultural learning in chimpanzees', *Public Library of Science One* **5**(5).

Horner, V. & Whiten, A. (2005), 'Causal knowledge and imitation/emulation switching in chimpanzees (pan troglodytes) and children (homo sapiens)', *Animal Cognition* **8**, 164–181.

Horner, V., Whiten, A., Flynn, E. & de Waal, F. B. M. (2006), 'Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children', *Proceedings of the National Academy of Sciences of the United States of America* **103**(37), 13878–13883.

Hudson-Kam, C. L. & Newport, E. (2005), 'Regularizing unpredictable variation: The roles of adult and child learners in language formation and change', *Language Learning and Development* **1**, 151–195.

Hudson-Kam, C. & Newport, E. L. (2009), 'Getting it right by getting it wrong: when learners change languages', *Cognitive Psychology* **59**, 30–66.

Hull, D. (2001), *Science and Selection: Essays on Biological Evolution and the Philosophy of Science*, Cambridge: Cambridge University Press.

Hurford, J. (1987), *Language and Number: the emergence of a cognitive system*, Oxford: Basil Blackwell.

Hurford, J. (1990), Nativist and functional explanations in language acquisition, *in* I. Roca, ed., 'Logical issues in language acquisition', Dordrecht: Foris, pp. 85–136.

Hurford, J. (2000), Social transmission favours linguistic generalisation, *in* J. Hurford, C. Knight & M. Studdert-Kennedy, eds, 'The evolutionary emergence of language: Social function and the orginis of linguistic forms', Cambridge: Cambridge University Press, pp. 324–352.

Hurford, J. (2002), Expression/induction models of language evolution: dimensions and issues, *in* T. Briscoe, ed., 'Linguistic evolution through language acquisition: Formal and computational models', Cambridge: Cambridge University Press, pp. 301–344.

Hurford, J. (2003), The language mosaic and its evolution, *in* M. H. Christiansen & S. Kirby, eds, 'Language Evolution: The States of the Art', Oxford: Oxford University Press.

Hurford, J. (2004), 'Human uniqueness, learned symbols and recursive thought', *European Review* **12**(4), 551–565.

Hurford, J. (2007), *The Origins of Meaning: Language in the light of evolution*, Oxford: Oxford University Press.

Hurford, J. (2011), *The Origins of Grammar: Language in the Light of Evolution*, Vol. Part I: Pre-Grammar, Oxford University Press.

Hurford, J., Studdert-Kennedy, M. & Knight, C., eds (1998), *Approaches to the Evolution of Language: Social and Cognitive Bases*, Cambridge: Cambridge University Press.

Ingram, D. (1989), *First Language Acquisition*, Cambridge: Cambridge University Press.

Isaac, G. (1978), 'Food sharing and human evolution: Archeological evidence from the plio-pleistocene of east africa', *Journal of Anthropological Research* **34**, 311–325.

Jablonka, E. & Lamb, M. J. (2005), *Evolution in Four Dimensions: genetic, epigenetic, behavioral, and symbolic variation in the history of life*, Cambridge, MA: MIT Press.

Jackendoff, R. (1996), 'How language helps us think', *Pragmatics and Cognition* **4**, 1–34.

Jacobs, R. C. & Campbell, D. T. (1961), 'The perpetuation of an arbitrary tradition through several generations of a laboratory microculture', *Journal of Abnormal and Social Psychology* **62**, 649–658.

Janik, V. & Slater, P. (2003), Traditions in mammalian and avian vocal communication, *in* D. Fragaszy & S. Perry, eds, 'The Biology of Traditions: Models and Evidence', Cambridge: Cambridge University Press, pp. 33–55.

Johnson, S. (2001), *Emergence*, London: Penguin Press.

Johnston, J. (1985), Cognitive prerequisites: the evidence from children learning english, *in* D. Slobin, ed., 'The Crosslinguistic Study of Language Acquisition', Vol. Vol 2, Hillsdale, NJ: Erlbaum, pp. 961–1004.

Jurafsky, D. & Martin, J. H. (2000), *Speech and Language Processing: An introduction to natural language processing, compuational linguistics, and speech recognition*, London: Prentice-Hall.

Kalish, M., Griffiths, T. & Lewandowsky, S. (2007), 'Iterated learning: Intergenerational knowledge transfer reveals inductive biases', *Psychonomic Bulletin and Review* **14**(2), 288–294.

Kameda, T. & Nakanishi, D. (2002), 'Cost-benefit analysis of social/cultural learning in a nonstationary uncertain environment', *Evolutionary Human Behavior* **23**, 373–393.

Kegl, J. (1994), 'The nicaraguan sign language: an overview', *Signpost* **7**, 24–31.

Keller, R. (1994), *On Language Change: The invisible hand in language*, London: Routledge.

Kessler, B. (2005), 'Phonetic comparison algorithms', *Transactions of the Philological Society* **103**(2), 243–260.

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

Kirby, S. (2000), Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners, *in* J. Hurford, M. Studdert-Kennedy & C. Knight, eds, 'Approaches to the evolution of language: Social and cognitive bases', Cambridge: Cambridge University Press, pp. 405–426.

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

Kirby, S. (2002*a*), Learning, botlenecks and the evolution of recursive syntax, *in* T. Briscoe, ed., 'Linguistic evolution through language acquisition: Formal and computational models', Cambridge: Cambridge University Press, pp. 173–204.

Kirby, S. (2002*b*), 'Natural language from artificial life', *Artificial Life* **8**(2), 185–215.

Kirby, S. (2007), The evolution of meaning-space structure through iterated learning, *in* C. Lyon, C. Nehaniv & A. Cangelosi, eds, 'Emergence of Coomunication and Language', Springer-Verlag, pp. 253–268.

Kirby, S., Cornish, H. & Smith, K. (2008), 'Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language', *Proceedings of the National Academy of Sciences of the United States of America* **105**, 10681–10686.

Kirby, S., Dowman, M. & Griffiths, T. (2007), 'Innateness and culture in the evolution of language', *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5241–5245.

Kirby, S. & Hurford, J. (1997), Learning, culture and evolution in the origin of linguistic constraints, *in* P. Husbands & I. Harvey, eds, 'ECAL97', Cambridge, MA: MIT Press, pp. 493–502.

Kirby, S. & Hurford, J. (2002), The emergence of linguistic structure: An overview of the iterated learning model, *in* A. Cangelosi & D. Parisi, eds, 'Simulating the evolution of language', Cambridge: Cambridge University Press, pp. 121–148.

Kirby, S., Smith, K. & Brighton, H. (2004), 'From ug to universals: linguistic adaptation through iterated learning', *Studies in Language* **28**(3), 587–607.

Kirby, S., Smith, K. & Cornish, H. (2008), Language, learning and cultural evolution: How linguistic transmission leads to cumulative adaptation, *in* R. Cooper & R. Kempson, eds, 'Language in Flux: Dialogue Coordination, Language Variation, Change and Evolution', London: College Publications, pp. 81–108.

Knowlton, B. J. & Squire, L. R. (1994), 'The information acquired during artificial grammar learning', *Journal of Experimental Psychology* **20**(1), 79–91.

Kroeber, A. L. & Kluckhohn, C. (1952), 'Culture: a critical review of the concepts and definitions', *Peabody Museum of American Archeology and Ethnology* **47**, 1–233.

Kupar, A. (2000), If memes are the answer, what is the question?, *in* R. Aunger, ed., 'Darwinizing Culture: The Status of Memetics as a Science', Oxford: Oxford University Press, pp. 175–188.

Laland, K. & Odling-Smee, F. J. (2000), The evolution of the meme, *in* R. Aunger, ed., 'Darwinizing Culture: The Status of Memetics as a Science', Oxford: Oxford University Press, pp. 121–142.

Laland, K., Richerson, P. & Boyd, R. (1993), Developing a theory of animal social learning, *in* C. Heyes & B. Galef, eds, 'Social Learning in Animals: The Roots of Culture', New York: Academic Press.

Laland, K. & Williams, K. (1997), 'Shoaling generates social learning of foraging information in guppies', *Animal Behavior* **53**, 1161–1169.

Landauer, T. K. & Dumais, S. T. (1997), 'A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge', *Psychological Review* **104**(2), 211–240.

Leakey, M. & Hay, R. (1979), 'Pliocene footprints in the laetolil beds at laetoli, northern tanzania', *Nature* **278**, 317–323.

Lee, N., Mikesell, L. & Joaquin, A. D. L. (2009), *The Interactional Instinct: the evolution and acquisition of language*, Oxford: Oxford University Press.

Lenneberg, E. H. (1967), *Biological Foundations of Language*, New York: Wiley.

Levenshtein, V. I. (1966), 'Binary codes capable of correcting deletions, insertions and reversals.', *Soviet Physics-Doklady* **10**, 707–710.

Lewin, R. (2005), *Human Evolution: An Illustrated Introduction*, 5th edn, Oxford: Blackwell.

Lewontin, R. C. (1974), *The genetic basis for evolutionary change*, New York: Columbia University Press.

Lieberman, P. (2007), 'The evolution of human speech: its anatomical and neural bases', *Current Anthropology* **48**(1), 39–66.

Line, P. (2010), Horizontal cultural transmisison of language: Investigating iterated learning with human dyads, Master's thesis, The University of Edinburgh.

Livingstone, D. & Fyfe, C. (1999), Modelling the evolution of linguistic diversity, *in* D. Floreano, J. Nicoud & F. Mondada, eds, 'ECAL99', Berlin: Springer-Verlag, pp. 704–708.

Locke, J. (1983), *Phonological Acquisition and Change*, New York: Academic Press.

Loewenstein, J. & Gentner, D. (2005), 'Relational language and the development of relational mapping', *Cognitive Psychology* **50**(4), 315–353.

Lumsden, C. J. & Wilson, E. O. (1981), *Genes, Mind, and Culture: The Coevolutionary Process*, Cambridge, MA: Harvard University Press.

Lund, K. & Burgess, C. (1996), 'Producing high-dimensional semantic spaces from lexical co-occurence', *Behavior, Research Methods, Instrumentation, and Computers* **28**, 203–208.

Lupyan, G. (2010), Beyond communication: Language modulates visual processing, *in* A. Smith, M. Schouwstra, B. de Boer & K. Smith, eds, 'The Evolution of Language: Proceedings of the 8th International Conference', pp. 443–444.

Lupyan, G., Rakison, D. & McClelland, J. (2007), 'Language is not just for talking: Redundant labels faciliate learning of novel categories', *Psychological Science* **18**(2), 1077–1083.

Luuk, E. & Luuk, H. (2010), 'The redundancy of recursion and infinity for natural language', *Cognitive Processing* .

MacLennan, B. & Burghardt, G. (1994), 'Synthetic ethology and the evolution of cooperative communication', *Adaptive Behavior* **2**, 161–187.

Macnamara, J. (1982), *Names for things: a study of human learning*, Cambridge, MA: MIT Press.

Macphail, E. (1982), *Brain and Intelligence in Vertebrates*, Oxford: Clarendom Press.

Mantel, N. (1967), 'The detection of disease clustering and a generalized regression approach', *Cancer Research* **27**(209-220).

Marcus, G. F. (1993), 'Negative evidence in language acquisition', *Cognition* **46**, 43–85.

Markman, E. M. & Wachtel, G. F. (1988), 'Children's use of mutual exclusivity to constrain the meaning of words', *Cognitive Psychology* **20**, 121–157.

Marler, P. (1957), 'Specific distinctiveness in the communication signals of birds', *Behaviour* **11**, 13–19.

Marler, P. & Tamura, M. (1962), 'Song "dialects" in three populations of white-crowned sparrow', *Condor* **64**, 368–377.

Matthews, C. (2009), The emergence of categorization: Language transmission in an iterated learning model using a continuous meaning space, Master's thesis, The University of Edinburgh.

Matthews, C., Kirby, S. & Cornish, H. (2010), The cultural evolution of language in a world of continuous meanings, *in* A. Smith, M. Schouwstra, B. de Boer & K. Smith, eds, 'The Evolution of Language: Proceedings of the 8th International Conference', pp. 451–452.

Maynard-Smith, J. & Szathmary, E. (2000), *The origins of life: From the birth of life to the origin of language*, Oxford: Oxford University Press.

McElreath, R., Lubell, M., Richerson, P., Waring, T., Baum, W. M., Edsten, E., Efferson, C. & Paciotti, B. (2005), 'Applying evolutionary models to the laboratory study of social learning', *Evolutionary Human Behavior* **26**, 483–508.

McGrew, W. (1992), *Chimpanzee Material Culture: Implications for Human Evolution*, Cambridge: Cambridge University Press.

Mellars, P. A. (2006), 'Going east: New genetic and archaeological perspectives on the modern human colonization of eurasia', *Science* **313**, 796–800.

Mesoudi, A. (2007), 'Using the methods of experimental social psychology to study cultural evolution', *Jornal of Social, Evolutionary, and Cultural Psychology* **1**(2), 35–58.

Mesoudi, A. (2008), 'Foresight in cultural evolution', *Biology and Philosophy* **23**(2), 243–255.

Mesoudi, A. & O'Brien, M. J. (2008), 'The cultural transmission of great basin projectile point technology i: An experimental simulation', *American Antiquity* **73**(1), 3–28.

Mesoudi, A. & Whiten, A. (2004), 'The hierarchical transformation of event knowledge in human cultural transmission', *Journal of Cognition and Culture* **4**(1), 1–24.

Mesoudi, A. & Whiten, A. (2008), 'The multiple roles of cultural transmission experiments in understanding human cultural evolution', *Philosophical Transactions of the Royal Society B* **363**, 3477–3488.

Mesoudi, A., Whiten, A. & Dunbar, R. (2006), 'A bias for social information in human cultural transmission', *British Journal of Psychology*, 405–423.

Mesoudi, A., Whiten, A. & Laland, K. (2004), 'Is human cultural evolution darwinian? evidence reviewed from the perspective of the origin of species', *Evolution* (1), 1–11.

Mesoudi, A., Whiten, A. & Laland, K. (2006), 'Towards a unified science of cultural evolution', *Behavioral and Brain Science*, 329–383.

Meteyard, L., Bahrami, B. & Vigliocco, G. (2007), 'Motion detection and motion verbs: Language affects low-level visual perception', *Psychological Science* (11), 1007–1013.

Miller, G. F. (2000), *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*, London: Doubleday.

Nakamura, M. (2000), 'Is human conversation more efficient than chimpanzee grooming? comparison of clique sizes', *Human Nature* (3), 281–297.

Nesse, R. M. & Williams, G. C. (1995), *Why we get sick: the new science of Darwinian medicine*, New York: Times Books.

Nettle, D. (1999), *Linguistic Diversity*, Oxford: Oxford University Press.

Nettle, D. (2009), *Evolution and Genetics for Psychology*, Oxford: Oxford University Press.

Newmeyer, F. J. (1991), 'Functional explanation in linguistics and the origins of language', *Language and Communication* (1-2), 3–28.

Newport, E. (1990), 'Maturational constraints on language learning', *Cognitive Science*, 11–28.

Niyogi, P. & Berwick, R. C. (1997), 'A dynamical systems model for language change', *Complex Systems*, 161–204.

Noble, J. (1999), 'Cooperation, conflict and the evolution of communication', *Adaptive Behavior*, 349–370.

Noble, W. & Davidson, I. (1996), *Human Evolution, Language and Mind*, Cambridge: Cambridge University Press.

Nowak, M. A., Komarova, N. L. & Niyogi, P. (2002), 'Computational and evolutionary aspects of language', *Nature*, 611–617.

Odling-Smee, F. J. & Laland, K. (2009), Cultural niche construction: evolution's cradle of language, *in* R. Botha & C. Knight, eds, 'The Pre-history of language', Oxford: Oxford University Press, pp. 99–122.

Odling-Smee, F. J., Laland, K. & Feldman, M. W. (2003), *Niche Construction: the neglected process in evolution*, Princeton University Press.

Oliphant, M. (1997), Formal approaches to innate and learned communication: Laying the foundation for language, PhD thesis, University of California, San Diego.

Pagel, M. & Mace, R. (2004), 'The cultural wealth of nations', *Nature* **428**, 275–278.

Parker, A. R. (2006), Evolving the narrow language faculty: was recursion the pivotal step?, *in* A. Cangelosi, A. Smith & K. Smith, eds, 'The Evolution of Language: Proceedings of the 6th International Conference', Singapore: World Scientific, pp. 239–246.

Petroski, H. (1992), *The evolution of useful things*, New York: Vintage Books.

Piattelli-Palmarini, M. (1989), 'Evolution, selection and cognition: from 'learning' to parameter setting in biology and in the study of language', *Cognition* **31**, 1–44.

Piattelli-Palmarini, M. & Uriagereka, J. (2004), The immune syntax: the evolution of the language virus, *in* L. Jenkins, ed., 'Variation and Universals in Biolinguistics', Oxford: Elsevier, pp. 341–377.

Pickering, M. J. & Garrod, S. (2004), 'Toward a mechanistic psychology of dialogue', *Behavioral and Brain Science* **27**, 169–225.

Pinker, S. (1994), *The Language Instinct: How the mind creates language*, New York: Harper-Collins.

Pinker, S. (1997), *How the Mind Works*, New York: W.W. Norton.

Pinker, S. (2003), Language as an adaptation to the cognitive niche, *in* M. H. Christiansen & S. Kirby, eds, 'Language Evolution: The States of the Art', Oxford: Oxford University Press, pp. 16–37.

Pinker, S. & Bloom, P. (1990), 'Natural language and natural selection', *Behavioral and Brain Science* **13**(4), 707–784.

Pinker, S. & Jackendoff, R. (2009), The components of language: what's specific to language, and what's specific to humans, *in* M. H. Christiansen, C. Collins

& S. Edelman, eds, 'Language Universals', Oxford: Oxford University Press, pp. 126–151.

Plummer, T. (2004), 'Flaked stones and old bones: Biological and cultural evolution at the dawn of technology', *American Journal of Physical Anthropology* **39**, 118–164.

Ptacek, M. (2000), 'The role of mating preferences in shaping interspecific divergence in mating signals in vertebrates', *Behavioral Processes* **51**, 111–134.

Pullum, G. & Scholz, B. (2002), 'Empirical assessment of stimulus poverty arguments', *The Linguistic Review* **19**, 9–50.

Quinn, M. (2001), Evolving communication without dedicated communication channels, *in* J. Keleman & P. Sosik, eds, 'Advances in Artificial Life: Proceedings of the 6th European Conference on Artificial Life', pp. 357–366.

Reber, A. S. (1969), 'Transfer of syntactic structures in synthetic languages', *Journal of Experimental Psychology* **81**, 115–119.

Rendell, L. & Whitehead, H. (2005), 'Culture in whales and dolphins', *Behavioral and Brain Sciences* **24**, 309–382.

Richerson, P. & Boyd, R. (2005), *Not by Genes Alone: How culture transformed human evolution*, Chicago: University of Chicago Press.

Ruhlen, M. (1994), *The origin of language: tracing the evolution of the Mother Tongue*, New York: Wiley.

Russell, S. & Norvig, P. (1995), *Artificial Intelligence: A Modern Approach*, London: Prentice-Hall.

Saffran, J., Aslin, R. & Newport, E. (1996), 'Statistical learning by 8-month olds', *Science* **274**, 1926–1928.

Sandler, W., Meir, I., Padden, C. & Aronoff, M. (2005), 'The emergence of grammar: Systematic structure in a new language', *Proceedings of the National Academy of Sciences of the United States of America* **102**(7), 2661–2665.

Schyns, P. G., Goldstone, R. L. & Thibaut, J.-P. (1998), 'The development of features in object concepts', *Behavioral and Brain Science* **21**, 1–54.

Scott-Phillips, T. C. & Kirby, S. (2010), 'Language evolution in the laboratory', *Trends in Cognitive Science* **14**(9), 411–417.

Scott-Phillips, T. C., Kirby, S. & Ritchie, G. (2009), 'Signalling signalhood and the emergence of communication', *Cognition* **113**(2), 226–233.

Selten, R. & Warglien, M. (2007), 'The emergence of simple languages in an experimental co-ordination game', *Proceedings of the National Academy of Sciences of the United States of America* **104**(18), 7361–7366.

Senghas, A. & Coppola, M. (2001), 'Children creating language: how nicaraguan sign language acquired a spatial grammar', *Psychological Science* **12**, 323–328.

Senghas, A., Kita, S. & Ozyurek, A. (2004), 'Children creating core properties of language: Evidence from an emerging sign language in nicaragua', *Science* **305**(5691), 1779–1782.

Shennan, S. J. & Steele, J. (1999), Cultural learning in hominids: a behavioural ecological approach, *in* H. Box & K. Gibson, eds, 'Mammalian Social Learning: Comparative and Ecological Perspectives', Cambridge: Cambridge University Press, pp. 367–388.

Shillcock, R., Kirby, S., McDonald, S. & Brew, C. (2001), Filled pauses and their status in the mental lexicon, *in* 'Proceedings of the 2001 conference on disfluency in spontaneous speech', Edinburgh, UK.

Skuse, D. (1984), 'Extreme deprivation in early childhood - ii. theorectical issues and a comparative review', *Journal of Child Psychology* **25**(4), 543–572.

Slobin, D. (1977), Language change in childhood and history, *in* J. Macnamara, ed., 'Language Learning and Thought', London: Academic Press, pp. 185–221.

Slobin, D. (1982), Universal and particular in the acquisition of language, *in* E. Wanner & L. Gleitman, eds, 'Language Acquisition: The State of the Art', Cambridge: Cambridge University Press.

Smith, J. M. & Szathmary, E. (1997), *The Major Transitions in Evolution*, Oxford: Oxford University Press.

Smith, K. (2003), Learning biases and language evolution, *in* S. Kirby, ed., 'Proceedings of Language Evolution and Computation Workshop/Course at ESSLLI', Vienna, pp. 22–31.

Smith, K. & Hurford, J. (2003), Language evolution in populations: extending the iterated learning model, *in* W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim & J. Ziegler, eds, 'Advances in Artificial Life - Proceedings of the European

Conference on Artificial Life (ECAL), Lecture Notes on Artificial Intelligence', Vol. 2801, Springer-Verlag, pp. 501–516.

Smith, K. & Kirby, S. (2008), 'Cultural evolution: implications for understanding the human langugae faculty and its evolution', *Philosophical Transactions of the Royal Society B* **363**, 3591–3603.

Smith, K., Kirby, S. & Brighton, H. (2003), 'Iterated learning: a framework for the emergence of language', *Artificial Life* **9**(4), 371–386.

Smith, K., Kirby, S. & Smith, A. (2008), 'The brain plus the cultural transmission mechanism determine the nature of language', *Behavioral and Brain Science* **31**(5), 533–534.

Smith, K. & Wonnacott, E. (2010), 'Eliminating unpredictable variation through iterated learning', *Cognition* **116**, 444–449.

Sokal, R. R. & Rohlf, F. J. (1995), *Biometry*, 3rd edn, New York: W.H. Freeman and Co.

Sperber, D. (1996), *Explaining Culture: a naturalistic approach*, Oxford: Blackwell.

Sperber, D. (2000), Why memes won't do, *in* R. Aunger, ed., 'Darwinizing Culture: The Status of Memetics as a Science', Oxford: Oxford University Press, pp. 163–174.

Steels, L. (2003), 'Evolving grounded communication for robots', *Trends in Cognitive Science* **7**(7), 308–312.

Stewart, D. (1858), *The collected works of Dugald Stewart: Vol. 10, Biographical memoirs of Adam Smith, William Robertson, Thomas Reid*, Edinburgh : T. Constable and Co.

Stone, L., Lurquin, P. F. & Cavalli-Sforza, L. L. (2007), *Genes, Culture, and Human Evolution*, Oxford: Basil Blackwell.

Tamariz, M. (2005), Exploring the Adaptive Structure of the Mental Lexicon, PhD thesis, The University of Edinburgh.

Tamariz, M. (2008), 'Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon', *Mental Lexicon* **3**(2), 259–278.

Tamariz, M. & Smith, A. (2008), Regularity in mappings between signals and meanings, *in* A. Smith, K. Smith & R. F. i Cancho, eds, 'The Evolution of

Language: Proceedings of the 7th International Conference (EVOLANG7)', pp. 315–322.

Teal, T. & Taylor, C. E. (2000), 'Effects of compression on language evolution', *Artificial Life* **6**(2), 129–143.

Tehrani, J. & Collard, M. (2002), 'Investigating cultural evolution through biologival phylogenetic analysis of turkmen textiles', *Journal of Anthropological Archeology* **21**(4), 443–463.

Theisen, C. A., Oberlander, J. & Kirby, S. (2010), 'Systematicity and arbitrariness in novel communication systems', *Interaction Studies* **11**(1), 14–32.

Tomasello, M. (1995), 'Language is not an instinct', *Cognitive Development* **10**, 131–156.

Tomasello, M. (1999), *The Cultural Origins of Human Cognition*, Harvard University Press.

Tomasello, M. (2004), 'What kind of evidence could refute the u.g. hypothesis?', *Studies in Language* **28**(3), 642–645.

Tomasello, M., Kruger, A. & Ratner, H. (1993), 'Cultural learning', *Behavioral and Brain Science* **16**, 495–552.

Tonkes, B. & Wiles, J. (2002), Methodological issues in simulating the emergence of language, *in* A. Wray, ed., 'The transition to language', Oxford: Oxford University Press.

Tooby, J. & Cosmides, L. (1992), The psychological foundations of culture, *in* J. Barkow, L. Cosmides & J. Tooby, eds, 'The adapted mind: Evolutionary psychology and the generation of culture', Oxford: Oxford University Press.

Tooby, J. & Cosmides, L. (2000), Toward mapping the evolved functional organization of mind and brain, *in* M. S. Gazzaniga, ed., 'The New Cognitive Neurosciences', Cambridge, MA: MIT Press, pp. 1167–1178.

Vogt, P. (2007), Group size effects on the emergence of compositional structures in language, *in* F. A. e Costa, L. M. Rocha, E. Costa & I. Harvey, eds, 'ECAL07', Springer-Verlag, pp. 405–414.

Vygotsky, L. S. (1962), *Thought and Language*, Cambridge, MA: MIT Press.

Watanabe, K. (2001), A review of 50 years of research on the japanese monkeys of koshima: status and dominance, *in* T. Matsuzama, ed., 'Primate Origins of Human Cognition and Behavior', Tokyo: Springer.

Watanabe, Y., Suzuki, R. & Arita, T. (2008), 'Language evolution and the baldwin effect', *Artificial Life and Robotics* **12**(1-2), 65–69.

Wexler, K. (1991), On the argument from the poverty of the stimulus, *in* A. Kasher, ed., 'In the Chomskyan Turn', Oxford: Blackwell.

White, R. (1982), 'Rethinking the middle/upper palaeolithic transition', *Current Anthropology* **23**, 169–192.

Whiten, A. (2005), 'The second inheritance system of chimpanzees and humans', *Nature* **437**, 52–55.

Whiten, A., Custance, D., Gomez, J.-C., Teixidor, P. & Bard, K. (1996), 'Imitative learning of artificial fruit processing in children (homo sapiens) and chimpanzees (pan trogloytes)', *Journal of Comparative Psychology* **110**(1), 3–14.

Whiten, A., Goodall, J., McGrew, W., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C., Wrangham, R. & Boesch, C. (1999), 'Culture in chimpanzees', *Nature* **399**, 682–685.

Whiten, A. & Mesoudi, A. (2008), 'Establishing an experimental science of culture: animal social diffusion experiments', *Philosophical Transactions of the Royal Society B* **363**, 3477–3488.

Wiese, H. (2004), *Numbers, language, and the human mind*, Cambridge: Cambridge University Press.

Winawer, J., Witthoft, N., Frank, M., Wu, L., Wade, A. & Boroditsky, L. (2007), 'Russian blues reveal effects of language on color discrimination', *Proceedings of the National Academy of Sciences of the United States of America* **104**(19), 7780–7785.

Winters, J. (2009), Adaptive structure, cultural transmission and language: investigating a population dynamic in human iterated learning, Master's thesis, The University of Edinburgh.

Worden, R. (1998), The evolution of language from social intelligence, *in* J. Hurford, M. Studdert-Kennedy & C. Knight, eds, 'Approaches to the evolution of language: Social and cognitive bases', Cambridge: Cambridge University Press, pp. 148–166.

Zahavi, A. & Zahavi, A. (1997), *The Handicap Priciple: a missing part of Darwin's Puzzle*, Oxford: Oxford University Press.

Ziman, J., ed. (2000), *Technological innovation as an evolutionary process*, Cambridge: Cambridge University Press.

Zuidema, W. (2003), How the poverty of the stimulus argument solves the poverty of the stimulus argument, *in* 'Advances in neural information processing systems 15 (Proceedings of NIPS'02)', Cambridge, MA: MIT Press, pp. 51–58.

Zuidema, W. (2005), The Major Transitions in the Evolution of Language, PhD thesis, The University of Edinburgh.

Appendix A

Instructions to participants in Experiments I-IV

Welcome to Alpha-3-6a in a galaxy far far away. We have encountered an intelligent alien life-form with its own form of language. You must try to learn this language as best you can.

Don't worry if you feel overwhelmed - the alien knows that this is a difficult task for you to master and it will do its best to understand everything that you say.

(Press ENTER to continue)

You will see a series of pictures and the way in which the alien would describe those pictures. Every now and then the alien will test your knowledge of the language by showing you a picture without any description. Simply write what you think the correct response is in the input box provided.

DON'T WORRY IF YOU FEEL YOU HAVE NOT YET MASTERED THE LANGUAGE!

The most important thing is to maintain good relations with the aliens and give it your best shot. ALWAYS GIVE AN ANSWER. That way the aliens will know you are trying. They will go out of their way to try to understand everything you say and they are very patient.

You will be given a break every 5 minutes or so.

If you have any questions please ask the experimentor now.

GOOD LUCK!

(press ENTER to start the tuition)

Appendix B1

Raw data from 4 transmission chains in Experiment I

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	hopa	vulepami	vulepami	vulepami	nepa	nepa	nepa	nepa	nepa	nepa	nepa
bounce	circle	blue	manehowu	nemine	vulepami	nepa	vulepami	nepa	nepa	nepa	nepa	nepa	nepa
bounce	circle	red	wuneho	nemine	lepa	vulepami	maho	maho	nepa	nepa	nepa	nepa	nepa
bounce	square	black	pamamapo	vulepami	vulepami	vulepami	nemene	nepa	nepa	nepa	nepa	nepa	nepa
bounce	square	blue	lemipo	veneme	lepa	vulepami	nepa	nepa	nepa	nepa	nepa	nepa	nepa
bounce	square	red	howu	nemine	nemene	maho	maho	maho	nepa	nepa	nepa	nepa	nepa
bounce	triangle	black	nehowu	pamapapo	vulepami	nemene	nemene	nepa	nepa	nepa	nepa	nepa	nepa
bounce	triangle	blue	nemi	mahole	nemi	vulepami	nepa	nepa	nepa	nepa	nepa	nepa	nepa
bounce	triangle	red	wunene	pali	nepa	vulepami	maho	nepa	nepa	nepa	nepa	nepa	nepa
horizontal	circle	black	lipapo	nepa	nepa	nepa	maho	nepa	nepa	nepa	nepa	nepa	nepa
horizontal	circle	blue	poliho	vemine	nemene	nepa	nepa	nepa	nepa	nepa	nepa	nepa	nepa
horizontal	circle	red	maho	maho	maho	maho	maho	maho	maho	nepa	nepa	nepa	nepa
horizontal	square	black	nehomami	pamapapo	pamapapo	maho	nepa	nepa	nepa	nepa	nepa	nepa	nepa
horizontal	square	blue	powuma	lemi	maho	maho	nepa	nepa	nepa	nepa	nepa	nepa	nepa
horizontal	square	red	wumaleli	maho	maho	maho	maho	nepa	nepa	nepa	nepa	nepa	nepa
horizontal	triangle	black	lilema	pamapapo	pamapapo	nepa	nepa	nepa	nepa	nepa	nepa	nepa	nepa
horizontal	triangle	blue	lemaho	nemi	nepa	maho	nepa	nepa	nepa	nepa	nepa	nepa	nepa
horizontal	triangle	red	lemilipo	maho	nepa	nemene	maho	nepa	nepa	nepa	nepa	nepa	nepa
spiral	circle	black	lepali	mapo	vulepami	pamano	pamano	pamano	pamano	nepa	nepa	nepa	nepa
spiral	circle	blue	lemi	nemene	wulepami	nemene	nepa	nepa	nepa	nepa	nepa	nepa	nepa
spiral	circle	red	nemine	lepa	nemine	maho	maho	maho	maho	nepa	nepa	nepa	nepa
spiral	square	black	pohomali	wulepami	vulepami	maho	nepa	nepa	nepa	nemene	nemene	nepa	nepa
spiral	square	blue	maholi	waheme	nemene	pamano	nemene	nemene	nemene	nemene	nemene	nemene	nemene
spiral	square	red	wupa	nemi	maho	maho	maho	nepa	nepa	nemene	nemene	nepa	nepa
spiral	triangle	black	wulepami	wulepami	maho	nepi	nemene	nemene	nepa	nepa	nepa	nepa	nepa
spiral	triangle	blue	nepa	nemu	nepa	nepi	vulepami	nepa	nepa	nepa	nepa	nepa	nepa
spiral	triangle	red	mahomine	nemine	lepa	maho	maho	maho	nepa	nepa	nepa	nepa	nepa

Table showing raw data for Chain A in Experiment I. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	lema	lehe	lehe	lehe	lehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe
bounce	circle	blue	manane	nawilehe	nawilehe	nawilehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe
bounce	circle	red	lehe	lehemu	nawilehe	nawilehe	lehe	lehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe
bounce	square	black	ninamahe	lehe	lehe	lehe	lehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe
bounce	square	blue	nawipuko	lehe	nawilehe	nawilehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe
bounce	square	red	puneniko	lehe	nawilehe	nawilehe	lehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe
bounce	triangle	black	maheko	lehe	lehe	lehe	lehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe
bounce	triangle	blue	hena	nawilehe	nawilehe	nawilehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe
bounce	triangle	red	leheni	lehe	lehe	lehe	lehe	lehe	ninalehe	ninalehe	ninalehe	ninalehe	ninalehe
horizontal	circle	black	konema	nepomu	koneko	mopuno	nina	winako	lehe	nina	lehe	nina	nina
horizontal	circle	blue	nepunani	manehe	mopulau	mopulau	nina	lehe	lehe	lehe	lehe	nina	nina
horizontal	circle	red	punema	koneko	mopunau	mopulau	ninakau	mopulau	mopulau	mopulau	nina	nina	nina
horizontal	square	black	naniwi	koneko	koneko	mopulau	nina	nina	nina	nina	nina	nina	nina
horizontal	square	blue	lemahene	punawi	koneko	mopulau	nina	nina	lehe	lehe	lehe	lehe	nina
horizontal	square	red	koneko	koneko	nekolau	mopulau	mopulau	lehe	lehe	lehe	nina	nina	nina
horizontal	triangle	black	puleni	nepomu	ninakoneko	nina	nina	lehe	lehe	lehe	nina	nina	nina
horizontal	triangle	blue	helewina	ninapomau	ninapolau	mopulau	nina	nina	nina	nina	nina	nina	nina
horizontal	triangle	red	koma	konekowi	ninakolau	ninakau	ninakau	nina	nina	nina	lehe	nina	nina
spiral	circle	black	winako	winako	winako	winako	winako	winako	wina	winako	wina	wina	wina
spiral	circle	blue	nawi	nepomuni	winapu	wina	wina	wina	wina	wina	wina	wina	wina
spiral	circle	red	wina	winamako	wina	wina	wina	wina	wina	wina	wina	wina	wina
spiral	square	black	wile	makoko	mopunu	wina	wina	wina	wina	winako	wina	wina	wina
spiral	square	blue	punawi	nepumehe	mopuno	wina	wina	wina	wina	wina	wina	wina	wina
spiral	square	red	lekopule	makomu	mopune	wina	wina	wina	wina	wina	wina	wina	wina
spiral	triangle	black	makoko	winako	winako	wina	wina	winako	winako	winako	winako	winako	winako
spiral	triangle	blue	malehewi	nawikok	wina	wina	wina	wina	wina	wina	wina	wina	winako
spiral	triangle	red	makopu	wina	wina	wina	wina	wina	wina	wina	wina	wina	winako

Table showing raw data for Chain B in Experiment I. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	kinimapi	kimei	hepini	tuge	tupim	tupim	minku	miniku	miniku	miniku	miniku
bounce	circle	blue	wikuki	miwn	miniku	tupim	tupim	tupim	miniku	miniku	miniku	miniku	miniku
bounce	circle	red	kikumi	miheniw	hepini	tupim	tupim	tupim	miniku	miniku	miniku	miniku	miniku
bounce	square	black	miwiniku	pemini	nige	miniku	mihunu	mihunu	miniku	miniku	tupim	tupim	tupim
bounce	square	blue	pinipi	kupini	tuge	tuge	tupim	tupim	tupin	miniku	tupim	tupim	tupim
bounce	square	red	kihemiwi	pon	mihenu	mihunu	miniku	tupim	tupim	tupim	tupim	tupim	tupim
bounce	triangle	black	miwimi	poi	poi	poi	poi	miniku	miniku	miniku	tupin	tupin	tupin
bounce	triangle	blue	nipi	mhip	mpo	tuge	miniku	tupim	tupin	tupin	tupin	tupin	tupin
bounce	triangle	red	wige	kuwpi	tupim	miniku	miniku	miniku	tupin	miniku	tupin	tupin	tupin
horizontal	circle	black	nihepi	mip	nige	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge
horizontal	circle	blue	wigemi	mpo	nige	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge
horizontal	circle	red	mahekuki	miniku	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge
horizontal	square	black	wimaku	nige	nige	mihenu	tuge	tuge	tuge	tuge	tuge	tuge	tuge
horizontal	square	blue	miniki	miniku	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge
horizontal	square	red	gepinini	poh	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge
horizontal	triangle	black	wikima	tuge	nige	nige	[null]	tuge	tuge	tuge	tuge	tuge	tuge
horizontal	triangle	blue	nipikuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge
horizontal	triangle	red	hema	weg	mpo	tuge	tuge	tuge	tuge	tuge	tuge	tuge	tuge
spiral	circle	black	pikuhemi	kuhepi	hepini	tupim	tupim	tupim	poi	poi	poi	poi	poi
spiral	circle	blue	kimaki	wige	tupim	tupim	tupim	tupim	poi	poi	poi	poi	poi
spiral	circle	red	pimikihe	mie	tupim	tupim	tupim	tupim	poi	poi	poi	poi	poi
spiral	square	black	gepihemi	hepinimi	hepini	mihenu	tupim	tupim	poi	poi	poi	poi	poi
spiral	square	blue	kunige	himini	miniku	tupim	tupim	tupim	tupin	poi	poi	poi	poi
spiral	square	red	miki	hipe	tupim	tupim	tupim	tupim	tupim	tupim	poi	poi	poi
spiral	triangle	black	mihe	pobo	nige	poi	poi	poi	poi	poi	poi	poi	poi
spiral	triangle	blue	winige	tupim	tupim	tupim	tupim	tupim	tupin	tupin	poi	poi	poi
spiral	triangle	red	kinimage	hipe	poi	tupim	tupim	tupim	tupim	poi	poi	poi	poi

Table showing raw data for Chain C in Experiment I. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	keni	wema	wepa	wepa	wema	wipe	wema	wema	wepa	wepa	wikepi
bounce	circle	blue	wumepihu	wape	wape	hukela	wema	wawakipi	wakepi	wema	wema	wepa	hidoku
bounce	circle	red	nihulu	nihulu	niveli	niveli	niveli	wepa	hudulu	wikepi	wikepi	wikepi	wikepi
bounce	square	black	pime	piwe	wepa	wepa	wipe	wuwu	wepa	wepa	wepa	wepa	wikepi
bounce	square	blue	memelu	hukile	nihulu	hekulu	heduku	hidulu	wakepi	wema	wepa	hidoku	hidoku
bounce	square	red	meluwa	nihulu	nihulu	niveli	wema	wawkipi	wema	wema	wikepi	wikepi	wikepi
bounce	triangle	black	wawapike	wawapike	wawapike	wipe	wema	wipe	wema	wema	wepa	wepa	wikepi
bounce	triangle	blue	wuhame	huwe	wipe	hekulu	hedulu	nihulu	wema	wikepi	wema	hidoku	hidoku
bounce	triangle	red	wani	pikewa	wawakipe	hekulu	wepa	wema	wema	wikepi	wikepi	wikepi	wikepi
horizontal	circle	black	wapiwu	pime	wipe	wipe	wipe	wepa	wipe	wepa	wepa	wipe	wipe
horizontal	circle	blue	mehuniha	niweli	wema	wema	wepa	nihulu	wepa	wepa	wipe	wepa	wipe
horizontal	circle	red	niluha	peluma	wepa	nirulu	wema	nihulu	wepa	wepa	wepa	wepa	wikepi
horizontal	square	black	kemepi	wume	nihulu	wipe	wipe	wipe	wipe	wipe	wipe	wipe	wepa
horizontal	square	blue	meni	pewa	hukela	nirulu	wema	wawakipi	wepa	wepa	wepa	wepa	wepa
horizontal	square	red	kepihuwu	wepamehu	hukela	nihulu	wepa	nihulu	wepa	wepa	wepa	wikepi	wepa
horizontal	triangle	black	piwu	pime	nihulu	wipe	wipe	jiduku	wipe	wepa	wepa	wipe	wipe
horizontal	triangle	blue	wuke	piwe	wema	wipe	wepa	wepa	wepa	wepa	wepa	wipe	wipe
horizontal	triangle	red	huhani	humepa	wawakipe	wepa	wema	nihulu	wepa	wepa	wepa	wipe	wikepi
spiral	circle	black	nimepa	piwe	wepa	wipe	wepa	wepa	hudulu	hidoku	hidoku	hidoku	hidoku
spiral	circle	blue	mepikelu	wuhili	huke;a	wepa	wepa	wepa	hidoku	hidoku	hidoku	hidoku	hidoku
spiral	circle	red	nimeni	lihuke	wipe	wipe	niveli	wepa	wakepi	wikepi	wikepi	hidoku	hidoku
spiral	square	black	lume	luwema	wepa	wepa	hedulu	hiduku	hudulu	hidoku	hidoku	hidoku	hidoku
spiral	square	blue	kewaha	hukela	wema	wakala	nihulu	wawakipe	hudulu	hidoku	hidoku	hidoku	hidoku
spiral	square	red	kewamewu	meka	hukela	hekulu	wawakipe	wawakipi	hudulu	hidoku	wikepi	wikepi	hidoku
spiral	triangle	black	luwaha	pehilu	wawakipe	wepa	wepa	hiduku	hidoku	hidoku	hidoku	hidoku	hidoku
spiral	triangle	blue	wapi	pikelu	wawakipe	wawakipe	nihulu	wpie	hidoku	hidoku	hidoku	hidoku	hidoku
spiral	triangle	red	humeni	nihuli	wema	wipe	wakepi	wipe	nihulu	wikepi	wikepi	wikepi	hidoku

Table showing raw data for Chain D in Experiment I. Shaded cells indicate that this item was selected to be seen by the next generation during training.

Appendix B2

Raw data from 4 transmission chains in Experiment II

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	huhunigu	niguki	wagukike	mukoni	muhapo	magini	nucapo	nucapo	nukapo	nucapo	nukapo
bounce	circle	blue	kemuniwa	kekoguni	mugike	mukoni	nucapo	mukapo	mukapo	mucapo	mukapo	nucapo	nucapo
bounce	circle	red	kihupo	hugukiki	nupokiki	mukapo	muhapo	mucapo	mucapo	mucapo	mucapo	nucapo	nucapo
bounce	square	black	wakiki	huwiku	mugike	mukoni	mutapo	nucapo	nucapo	nucapo	nucapo	nucapo	nucapo
bounce	square	blue	pokikehu	nugukike	nugikinu	koni	mukapo	mugini	mukapo	mucapo	nukapo	mucapo	nucapo
bounce	square	red	waguhuki	muguki	wekike	mukapo	muckapo	mucapo	mukapo	mukapo	nucapo	mucapo	nukapo
bounce	triangle	black	nihu	wakiki	koni	koni	mugeni	mugenini	nucapo	nucapo	nukapo	nucapo	nucapo
bounce	triangle	blue	niguki	wukeki	koni	koni	mukapo	mucapo	mukapo	mukapo	nucapo	nukapo	nukapo
bounce	triangle	red	koni	koni	mukoni	mukoni	muhapo	nucapo	mucapo	mukapo	nucapo	mucapo	nucapo
horizontal	circle	black	muwapo	muguki	wapo	kapo	kapo	hapo	hapo	hapo	kapo	hapo	hapo
horizontal	circle	blue	powa	wapo	nuha	hapo	hapo	hapo	hapo	hapo	kapo	hapo	hapo
horizontal	circle	red	hukinimu	niguki	hapo	kapo	kapo	kapo	kapo	kapo	kapo	hapo	hapo
horizontal	square	black	wako	muwapo	mukeki	nugeki	kapo	hapo	hapo	hapo	kapo	hapo	hapo
horizontal	square	blue	hukeko	waku	kapo	hapo	hapo	hapo	hapo	hapo	hapo	hapo	hapo
horizontal	square	red	pohumu	gukike	kapo	huni	hapo	kapo	kapo	kapo	kapo	hapo	hapo
horizontal	triangle	black	muko	nihu	huni	kapo	kini	hapo	hapo	hapo	hapo	hapo	hapo
horizontal	triangle	blue	kokeguke	nihu	koni	koni	hapo	hapo	hapo	kapo	hapo	hapo	hapo
horizontal	triangle	red	kimu	wakeke	huni	kapo	kapo	kapo	kapo	kapo	kapo	hapo	hapo
spiral	circle	black	kekewa	wakeko	nugeke	kapo	wakeke	wunigni	maginini	waginini	wagnini	wagini	waginini
spiral	circle	blue	komuhuke	wakigu	huguni	wakeke	wakeke	nugini	wagini	wagini	wagini	wagini	waginini
spiral	circle	red	kopo	kopo	nugoni	nugikini	nugikini	mugingi	nugikini	nugokini	nugakini	wagini	waginini
spiral	square	black	huwa	wakuki	wakeke	wakeki	mugenini	wagigini	nugini	wagini	waginini	waginini	waginini
spiral	square	blue	hukike	huguni	kapo	noguni	wakeke	wagini	wagini	wagini	waginini	waginini	wagini
spiral	square	red	ponikiko	nuguki	mukapo	wakeke	nugikini	mugini	wagini	wagini	nugakini	waginini	waginini
spiral	triangle	black	kowagu	guni	wakiki	wakiki	wakeni	mugini	magini	waginini	wagini	nuakini	macini
spiral	triangle	blue	kokihuko	muguni	hapo	nuguni	mugeni	mugini	wagini	wagini	waginini	nuakini	waginini
spiral	triangle	red	kiwanike	nuguki	wakeki	nugeni	wakeke	wagini	wagini	nugakini	nugokini	nuakini	nakaini

Table showing raw data for Chain A in Experiment II. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	kuwo	nahenko	lugigipi	gulo	gulo	naheku	pulo	pulo	pulo	pulo	pulo
bounce	circle	blue	wonagi	huwo	lugigipi	legigipi	legigipi	lugipigi	gulo	neheku	nahepu	nehepu	nepu
bounce	circle	red	pelu	gulo	gulo	gulo	naheku	naheku	pulo	pulo	nagigipi	nagigipi	pili
bounce	square	black	wogipena	huko	naheku	lugigipi	lugigipi	naheku	pulo	pulo	pulo	pulo	nagigihi
bounce	square	blue	napena	huko	henko	naheku	legigipi	legigipi	naheki	neheki	naheki	nehepu	pulo
bounce	square	red	penapiku	penapiku	naheku	naheku	naheku	legigipi	pili	pili	pili	pili	nepi
bounce	triangle	black	gapinahe	naheku	gunko	gunko	nagigeki	gunko	pulo	pulo	pulo	pulo	nepi
bounce	triangle	blue	hewoku	luwenko	naheku	henko	nageku	legigipi	pulo	neheki	naheki	nehepi	nepu
bounce	triangle	red	giku	giko	naheku	naheku	nageku	nekigeki	pili	pili	pili	pili	pili
horizontal	circle	black	lugigipi	lugigipi	pelu	gulo	gulo	gulo	naheku	nagegepi	pulo	nepu	pulo
horizontal	circle	blue	naheku	ligigipi	gulo	gulo	gulo	pilu	nagegepi	hepu	nepu	nepu	pulo
horizontal	circle	red	kuluwo	naheku	gulo	gulo	gulo	pihu	nagigipi	pulo	pili	nepu	pulo
horizontal	square	black	wogiluku	lugowo	pelu	pilu	pilu	pilu	nagigipe	pulo	pulo	nepu	pili
horizontal	square	blue	gikuna	pihu	pihu	pihu	pilu	pihu	pili	nehepu	nehi	napu	nepi
horizontal	square	red	napeheku	guko	neku	pihu	pihu	pihu	gulo	nagigipi	pili	pili	pili
horizontal	triangle	black	penalu	giku	pelu	pilu	pilu	gulo	hepu	nepu	nepu	nepu	nepi
horizontal	triangle	blue	pihena	pihu	pihu	pilu	pilu	pihu	hipu	nepu	nepu	nepu	nepu
horizontal	triangle	red	naku	naku	pilu	pihu	pihu	pihu	nagigipi	pili	pili	pili	pili
spiral	circle	black	lugana	galu	nekigeki	nekigeki	lugigipi	gulo	nagigipi	pulo	pulo	pulo	pulo
spiral	circle	blue	heku	henku	heki	legigipi	gulo	nekigeki	nagigipi	nehepu	nahepu	naheki	nepi
spiral	circle	red	wonalupe	lugibi	legigipi	naheku	gulo	legigipi	nagigipi	nagigipi	nagegepi	nagigipi	pulo
spiral	square	black	galukuna	wugo	gelu	lugigipi	lugigipi	naheku	naheku	nagegepi	pulo	pulo	pulo
spiral	square	blue	napiwo	naheku	lugigipi	legigipi	legigipi	naheku	nahepu	nagegepi	nahepu	nehepu	nagighi
spiral	square	red	lupiwo	naheku	lugigipi	lugigipi	lugigipi	legigipi	pili	nagigipi	nagigipi	nagegepi	pulo
spiral	triangle	black	nahe	pelu	geki	henko	lugigipi	legigipi	naheku	neheku	pulo	pulo	pulo
spiral	triangle	blue	pihe	lungo	naheku	lugigipi	legigipi	legigipi	pulo	nehepu	neheki	nehepi	nepi
spiral	triangle	red	galu	nahenko	heki	heku	lugigipi	lugigipi	pulo	nagigipi	nagigipi	nagegepi	pulo

Table showing raw data for Chain B in Experiment II. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	kalu	humoneki	humonpiki	luneki	manolaki	trilaki	manolaki	manolaki	manolaki	manolaki	manolaki
bounce	circle	blue	mola	kahupiki	kumopiki	nane	lunaki	lunalaki	lunalaki	lunlaki	lunolaki	lunolaki	lunolaki
bounce	circle	red	pihukimo	nemalo	nane	humoneki	lunaki	trilaki	manolaki	manolaki	humolaki	humolaki	humolaki
bounce	square	black	moki	kunapiki	humonpiki	nano	mano	manolaki	manolaki	manolaki	trilaki	trilaki	manolaki
bounce	square	blue	luneki	kaneki	humopiki	nuleki	manolaki	manolaki	manolaki	humalaki	manolaki	manolaki	lunalaki
bounce	square	red	lanepi	lahupino	kumonaki	trileki	trilaki	manolaki	manolaki	humalaki	humalaki	humalaki	humalaki
bounce	triangle	black	nane	nane	humoneki	huleki	mulaki	trilaki	trilaki	trilaki	trilaki	trilaki	trilaki
bounce	triangle	blue	kalakihu	hokune	humopiki	trileki	trilaki	lunlaki	lunlaki	trilaki	trilaki	trilaki	trilaki
bounce	triangle	red	mokihuna	naki	nane	luneki	trilaki	trilaki	trilaki	trilaki	trilaki	trilaki	trilaki
horizontal	circle	black	nelu	maneki	malo	mano	mano	mano	mano	mano	mano	mano	mano
horizontal	circle	blue	kanehu	malo	korane	naleki	lunaki	manolaki	manolaki	mano	mano	mano	mano
horizontal	circle	red	namopihu	kuneki	luneki	luniki	humalaki	mano	mano	mano	mano	mano	mano
horizontal	square	black	lumonamo	huneki	humano	mano	mano	mano	mano	mano	mano	mano	mano
horizontal	square	blue	kinehune	humonamo	nari	maleki	mulaki	manolaki	manolaki	mano	mano	mano	mano
horizontal	square	red	lahupine	kahune	kuneki	naleki	mano	mano	manolaki	manolaki	manolaki	manolaki	manolaki
horizontal	triangle	black	kapihu	malo	humona	mano	mano	mano	mano	mano	mano	mano	mano
horizontal	triangle	blue	humo	humo	humo	keleki	manolaki	mano	mano	mano	mano	mano	mano
horizontal	triangle	red	lahupiki	luneki	luneki	muleki	trilaki	mano	manolaki	mano	mano	mano	mano
spiral	circle	black	pilu	malo	kuneki	naleki	mano	mano	mano	mano	mano	mano	manolaki
spiral	circle	blue	neki	kahune	pilu	luneki	lunaki	manolaki	manolaki	manolaki	manolaki	manolaki	manolaki
spiral	circle	red	pinemohu	luneki	luneki	nane	humalaki	manolaki	manolaki	humalaki	manolaki	manolaki	manolaki
spiral	square	black	kilamo	pilu	pilu	mano	mano	mano	mano	manolaki	manolaki	manolaki	manolaki
spiral	square	blue	kahuki	pilu	pilu	luneki	humalki	manolaki	manolaki	manolaki	manolaki	manolaki	manolaki
spiral	square	red	neluka	namupiku	kuneki	meneki	mano	humalaki	humalaki	humalaki	manolaki	manolaki	manolaki
spiral	triangle	black	luki	luneki	kuneki	mano	mano	mano	manolaki	manolaki	manolaki	manolaki	manolaki
spiral	triangle	blue	namola	kuneki	pilu	luneki	trilaki	manolaki	manolaki	manolaki	manolaki	manolaki	trilaki
spiral	triangle	red	lumoka	haneki	luneki	maleki	kelaki	humalaki	humalaki	manolaki	manolaki	manolaki	manolaki

Table showing raw data for Chain C in Experiment II. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	melipa	mawehika	meme	liga	liga	liga	liga	liga	liga	liga	liga
bounce	circle	blue	pamu	pamu	paweliga	hinoliki	hinoliki	paweli	paweli	paweli	paweli	paweli	paweli
bounce	circle	red	mewega	linuhiko	memenu	hinoliki	meme	meme	meme	meme	meme	meme	meme
bounce	square	black	gamuwe	maweliga	liga	liga	liga	liga	liga	liga	liga	liga	liga
bounce	square	blue	linuhiko	luhiko	hinoliki	hinoliki	hinoliki	hinoliki	hinoliki	hinoliki	paweli	paweli	paweli
bounce	square	red	komehi	hikomeli	highili	meme	meme	meme	meme	meme	meme	meme	meme
bounce	triangle	black	hiko	hiko	hiko	highili	liga	liga	liga	liga	liga	liga	liga
bounce	triangle	blue	palime	palime	pawemeli	hinoliki	hinoliki	hinoliki	hinoliki	hinoliki	paweli	paweli	paweli
bounce	triangle	red	gawe	gawe	hiko	meme	meme	meme	meme	meme	meme	meme	meme
horizontal	circle	black	hiwenuko	linu	meme	meme	memenu	liga	liga	liga	liga	liga	liga
horizontal	circle	blue	nuhiwenu	liga	meme	menu	paweli	memenu	memenu	memenu	menenu	memenu	memenu
horizontal	circle	red	memenu	memenu	liga	meme	meme	meme	meme	meme	meme	memenu	memenu
horizontal	square	black	paweko	liga	liga	menu	meme	liga	liga	liga	liga	liga	liga
horizontal	square	blue	konulipa	pawehiko	meme	liga	memenu	memenu	memenu	memenu	honolike	honolike	memenu
horizontal	square	red	linu	menu	menu	meme	meme	meme	meme	meme	meme	meme	meme
horizontal	triangle	black	mume	meme	hiko	liga	liga	liga	liga	liga	liga	liga	liga
horizontal	triangle	blue	pawemeli	mume	liga	menu	paweli	paweli	paweli	memenu	menenu	memenu	memenu
horizontal	triangle	red	liga	liga	menu	paweliga	meme	meme	meme	meme	meme	meme	meme
spiral	circle	black	melime	melime	meme	hinoliki	liga	liga	liga	liga	liga	liga	liga
spiral	circle	blue	munuko	pawemeli	paweli	meme	paweli	paweli	paweli	paweli	paweli	paweli	paweli
spiral	circle	red	komume	hikoliga	paweli	paweli	meme	meme	meme	meme	meme	meme	meme
spiral	square	black	numekopa	mumehiko	liga	meme	paweli	liga	liga	liga	liga	liga	liga
spiral	square	blue	wega	notomeli	paweliga	paweli	paweli	paweli	paweli	paweli	paweli	paweli	paweli
spiral	square	red	higahili	highili	paweli	meme	meme	meme	meme	meme	meme	meme	meme
spiral	triangle	black	pawe	pawemali	hiko	memenu	liga	liga	liga	liga	liga	liga	liga
spiral	triangle	blue	gahi	paweliga	paweli	paweli	paweli	paweli	paweli	paweli	honolike	honolike	honolike
spiral	triangle	red	muwemeko	pawe	memenu	memenu	meme	meme	meme	meme	meme	meme	meme

Table showing raw data for Chain D in Experiment II. Shaded cells indicate that this item was selected to be seen by the next generation during training.

Appendix B3

Raw data from 4 transmission chains in Experiment III

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	kalu	lanapi	keluupilu	kahona	kahona	neki	nekiplu	nekinono	nerena	nehoplo	reneplo
bounce	circle	blue	mola	kalu	kanupilu	lehona	kahuna	laneki	lapipilu	lapklu	leneho	lahoplo	lareplo
bounce	circle	red	pihukimo	kalu	napilu	kahona	kanana	kanana	kanana	renana	renana	rehoplo	reneplo
bounce	square	black	moki	lanapi	lapilu	kanupilu	kahona	kahopilu	nekipilu	nekeno	nekeno	nereplo	nereplo
bounce	square	blue	luneki	lapalu	lunahoma	lehona	lanuna	lanpilu	lanepilu	lahoki	kapilu	laneplo	laneplo
bounce	square	red	lanepi	kanepi	luhona	lanpilu	kahuna	kahepilu	kaneki	nekipilu	renana	replo	reneplo
bounce	triangle	black	nane	kilahuna	kahoma	kanupilu	lunona	lanpilu	nekinono	nekiplu	nekiplo	nekiplo	nekiplo
bounce	triangle	blue	kalakihu	lamuna	kepihoma	kanupilu	kahuna	nehoki	lapiranana	kanana	lepilo	lakiplo	lakiplo
bounce	triangle	red	mokihuna	pinamula	nepalu	lapilu	nanuna	kahopilu	kapilu	kanana	rekiplo	rahoplo	rekiplo
horizontal	circle	black	nelu	napilu	pilu	neki	kahoneki	nepilu	nekepilu	nekeno	nereki	neheki	faneki
horizontal	circle	blue	kanehu	pilu	pilu	lanike	kaneki	lanepilu	lahoki	laneki	laneki	lahoki	lareki
horizontal	circle	red	namopihu	pilu	kanupilu	kaneki	kanneki	kane	kaponeki	reneki	renato	reneki	reneki
horizontal	square	black	lumonamo	pilu	laneki	neneki	neki	neki	nepilu	naneki	nereki	nereki	nereki
horizontal	square	blue	kinehune	nahuna	kaneki	laneki	laneki	laneki	laneki	lanoki	lanena	lereki	laneki
horizontal	square	red	lahupine	humo	kaneki	kaneki	kaneki	kaneki	kaneki	luni	renana	renana	renana
horizontal	triangle	black	kapihu	kahumo	neki	neki	luneki	nekipilu	nekeni	keniko	nekeki	nekeki	lakaki
horizontal	triangle	blue	humo	neki	homa	neki	kaneki	lanpilu	lapineki	laneki	laneki	lakeki	lakiki
horizontal	triangle	red	lahupiki	pilu	kaneki	naneki	naneki	kenepilu	kaphiki	reneki	raneki	raheki	rekiki
spiral	circle	black	pilu	kinepilu	pilu	pilu	kahopilu	nekopilu	nepipilu	nahokilu	nepilu	nehopilu	renepilu
spiral	circle	blue	neki	kinepila	lepilu	lepilu	kapilu	lahopilu	lahopilu	lahopilu	lehopilo	lahopilu	larepilu
spiral	circle	red	pinemohu	lamuna	napilu	kanpilu	kanpilu	kahopilu	kapilu	rehopilu	rehopilu	rehopilu	rehepilu
spiral	square	black	kilamo	kahuna	kahona	kapilu	kapilu	nekilu	nekpilu	kekilu	nehopilu	nepilu	nerepilu
spiral	square	blue	kahuki	luneki	luneki	lanpilu	lanpilu	lanepilu	lanepilu	lanpilu	lanpilo	lanepilu	lanepilu
spiral	square	red	neluka	lanuka	napilu	kahona	kapilu	kahopilu	kanepilu	kanpilu	rehopilu	repilu	renepilu
spiral	triangle	black	luki	kalu	kinu	lupilu	lupilu	nekipilu	nepilu	nepilu	nepilu	nekipilu	lakipilu
spiral	triangle	blue	namola	neki	nakemi	lepilu	kapilu	lanepilu	lapipilu	lakipilu	lakipilo	lakipilu	lakipilu
spiral	triangle	red	lumoka	napulu	kaneki	nepilu	napilu	kapilu	kapipilu	rekepilu	rakipilu	rahopilu	rekipilu

Table showing raw data for Chain A in Experiment III. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	melipa	mewnihi	mewena	mewega	menowa	helahilhil	meknowa	meknowa	mehnoa	mehnoah	mekoah
bounce	circle	blue	pamu	gamewe	mowoga	mewega	menowa	helhilhil	menoah	meknoa	mehknoah	meknoa	mekhoah
bounce	circle	red	mewega	owumuga	mewega	pewega	menowa	meknowna	meknoah	meknoa	mehknoah	mekhnoah	mehkoa
bounce	square	black	gamuwe	muwenega	mowoga	monowa	menowa	meena	hilahillhil	meknowa	mehnoa	mehnoah	mehkoa
bounce	square	blue	linuhiko	mowoga	mowoga	monowa	menowna	menowa	menahilhil	meknoah	meknoa	mekhnoa	mekoah
bounce	square	red	komehi	pawenego	mowoga	monowa	menowna	pewega	menahilhil	meknoah	mehknoah	mehnoa	mehkoah
bounce	triangle	black	hiko	mowenghi	mowoga	kewona	kenowa	mena	mehnoha	mehnoha	mehnoah	mehnoah	meknoah
bounce	triangle	blue	palime	palinia	mowoga	meena	menowa	meknowa	meknowah	meknoah	meknoa	meknoah	meknoah
bounce	triangle	red	gawe	mewenega	mewoga	kewona	menowa	meena	menaoh	meknoa	mehknoah	meknoa	mekhoa
horizontal	circle	black	hiwenuko	mewnahi	lina	mewega	menowa	pewega	mena	mena	pega	pega	mekhoah
horizontal	circle	blue	nuhiwenu	menunana	lina	pewega	pegewa	pewega	pega	pega	menu	menu	menu
horizontal	circle	red	memenu	liga	pewega	mewega	menowna	perega	palin	palim	palim	palim	mekoa
horizontal	square	black	paweko	palin	palin	palin	kenowa	meena	mena	menu	pega	palim	mekoa
horizontal	square	blue	konulipa	peenla	palin	palin	pewega	pegawa	meknoa	meknoa	menu	pega	meknoah
horizontal	square	red	linu	lega	palin	palin	palin	palin	palin	palim	pegu	pega	mekhoah
horizontal	triangle	black	mume	meewena	meweena	lina	kenowna	perega	pegu	pegu	pegu	pega	mekoah
horizontal	triangle	blue	pawemeli	lina	lina	lina	kenowna	pewage	mena	pegu	pega	pega	mehkoa
horizontal	triangle	red	liga	lega	lina	lina	kenowna	perega	pegas	pegas	palim	palim	palim
spiral	circle	black	melime	memilhi	meena	helahilhil	kenowna	hellahilhil	hellahillhill	helahillhil	hellahilhil	helahillhil	meknoa
spiral	circle	blue	munuko	memeena	helhilhil	helahilhil	helahilhil	helhilhil	hellahillhill	hellahillhill	hellahilhil	hellahillhil	helihilhill
spiral	circle	red	komume	helhilhil	helahilhil	meena	meena	helhilhil	helahilhil	hellahilhil	hellahilhil	helahillhill	helihillhil
spiral	square	black	numekopa	pawethi	helhilhil	helhilhil	helahilhil	hellahilhil	hellahillhill	helahilhil	helahilhil	helahillhill	mekoah
spiral	square	blue	wega	helahilhil	kewona	helhilhil	helhilhil	knowna	helahilhill	helahilhill	helahillhill	hellahilhil	mekhoah
spiral	square	red	higahili	kenowma	helahilhil	helhilhil	meena	hilhilhil	helahilhil	helahilhill	helahilhill	helahilhill	helihillhill
spiral	triangle	black	pamu	hilihihi	kenowna	kenowna	kenowna	hellahillhill	helahilhill	hellahilhil	hellahilhil	helahilhill	mekhoah
spiral	triangle	blue	gahi	peneka	peneena	kenowna	helahilhil	helhilhil	helahillhill	helahilhil	helahilhil	hellahillhill	mehkoa
spiral	triangle	red	muwemeko	pawena	meena	kenowna	helahilhil	hilhilhil	hellahillhill	hellahilhil	meknoah	meknoah	mekhoa

Table showing raw data for Chain B in Experiment III. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	kuwo	wogilupu	wogilopa	wogipenal	penalowgi	wogipenal	wogipenal	wogininalgi	wonilunalgi	wolilunagi	wolilunagi
bounce	circle	blue	wonagi	napiwo	wogilupe	wogipenal	woginepal	woginepal	wogipenal	wogipenalgi	wonilunalgi	wogilenani	wolilunagi
bounce	circle	red	pelu	wogilupu	wogilupa	nepalwogi	wogilugan	wogiluna	wogilunagi	wogilunagi	woginunagi	wolilunani	wonunali
bounce	square	black	wogipena	wogipena	nepalwogi	woginepal	wogipenal	wogipenalgi	wogipenalgi	penlunagi	woninagi	wolinunalgi	wonunali
bounce	square	blue	napena	nape	wogilope	nepalwogi	woginepal	woginepal	wogipenalgi	wogipenalgi	wolinalgi	wolinunalgi	wolinunagi
bounce	square	red	penapiku	heka	penalogi	wogiluna	wogiluna	wogilunowgi	wogipenalgi	penaninagi	wogipenalgi	wolinulalgi	wolinunali
bounce	triangle	black	gapinahe	wogipena	wogipenal	wogipenal	wogipenal	wogipenal	woginepalgi	woginunagi	wonalgi	wolilenul	wolilunali
bounce	triangle	blue	hewoku	heka	wogilupe	wogipenal	nepalowgi	wogineptune	woginepalgi	nulagi	wolinalgi	wolinulagi	wolinunali
bounce	triangle	red	giku	lugana	wogipule	wogipenal	luganowgi	wogilunagi	woginepal	nenalgi	wogiwenagi	wolinugi	wolinunali
horizontal	circle	black	lugigipi	kuwo	lugana	wogiluna	penalowgi	penalgi	nunagi	wonagi	wogilunagi	penagi	penali
horizontal	circle	blue	naheku	napena	lugana	penalike	nepalowgi	nunapagi	nunagi	woginalgi	wolilunagi	penagi	penagi
horizontal	circle	red	kuluwo	heka	lugana	nepalowgi	luganowgi	lunagi	lunagi	woginal	wogilunagi	penalgi	penalgi
horizontal	square	black	wogiluku	lugana	lugana	penalwogi	penalowgi	penalgow	penalgi	penalgi	nulagi	penalgi	penagi
horizontal	square	blue	gikuna	napena	lugana	penalike	lugana	nepalgi	penalgi	penalgi	wonalgi	penalgi	penagi
horizontal	square	red	napeheku	napiwo	penalike	lunawogi	lugana	lunagi	penalgi	penalgi	penalgi	penalgi	penalgi
horizontal	triangle	black	penalu	lugana	penalowgi	penalowgi	nepalike	penalgi	nepalgi	nulagi	nunalgi	nulani	penali
horizontal	triangle	blue	pihena	penaliku	lugana	lugana	penalowgi	neptungi	nepalgi	nunagi	wonagi	penal	penal
horizontal	triangle	red	naku	giku	lugana	wogipenal	lugana	lungagi	nepalgi	nulagi	wogalgi	penul	penali
spiral	circle	black	lugana	kuwo	reki	reki	pike	pike	like	wonagi	wonagi	wonagi	wonalgi
spiral	circle	blue	heku	napena	heki	heki	nepalike	nike	penal	wonulgi	wolilunagi	wonagi	wonagi
spiral	circle	red	wonalupe	heka	heki	heki	like	like	wogiluna	woginal	nolunagi	wolinagi	wonalgi
spiral	square	black	galukuna	wogipena	reki	reki	hike	pike	penal	penal	nenalgi	wonalgi	wonalgi
spiral	square	blue	napiwo	napena	kibve	neki	nike	nike	penal	penal	wolinulagi	wonalgi	wonagi
spiral	square	red	lupiwo	wogilupe	neki	heki	like	like	penal	penal	penal	wonalgi	wonalgi
spiral	triangle	black	nahe	kuwo	kipe	kipe	pike	pike	nike	nike	nike	wonali	wonali
spiral	triangle	blue	pihe	giku	neki	kipe	nepalike	nike	like	nike	wonalgi	wolinag	wonagi
spiral	triangle	red	galu	giku	neki	heki	hike	like	nepal	nike	woninal	wolinal	wonalgi

Table showing raw data for Chain C in Experiment III. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	huhunigu	pikoku	wikiko	wikiko	winekuki	winekuki	winikike	winikiko	wunkiko	winikiko	winikiko
bounce	circle	blue	kemuniwa	huniki	hukiki	kunkuki	kunkuki	hunekuki	honekiko	honekiko	kunkike	hunekiko	hunekiko
bounce	circle	red	kihupo	piko	pokiko	ponekuki	ponekuki	ponekuki	punekiko	ponekiko	punkiko	punekiko	punekiko
bounce	square	black	wakiki	wukiki	winekiko	winikiko	winukuki	winikuki	winikeko	winikiko	winikiko	winikiko	winikiko
bounce	square	blue	pokikehu	ponuko	kunikeko	hunekuki	hunekuki	hunekuki	kunekiko	kunekiko	ponekiko	hunekiko	hunekiko
bounce	square	red	waguhuki	poku	ponekiko	ponekuki	punekuki	punikuki	ponekiko	punekiko	pinkiko	punkiko	hunekiko
bounce	triangle	black	nihu	kikiki	kikiki	winekiko	wikekuki	wanikuki	winikiko	winikiko	winekiko	winekiko	punikiko
bounce	triangle	blue	niguki	hukeko	hukiki	kunekuki	kunekuki	kunikuki	kunekiko	kikekiko	pinekiko	ponikiko	winikike
bounce	triangle	red	koni	koni	ponekiko	ponekiko	ponekuki	punekuki	punekiko	punikiko	punkiko	punkiko	punkiko
horizontal	circle	black	muwapo	wuniki	wineko	wineko	wineko	wineke	wineke	wineke	wuneke	winekike	punike
horizontal	circle	blue	powa	pinokiki	huneko	kuneko	kuneko	kunike	honeke	honeke	kineke	hunike	wineke
horizontal	circle	red	hukinimu	kuniko	ponukeko	poneko	poneko	ponike	punike	ponike	puneke	punike	winikike
horizontal	square	black	wako	wako	wikeko	wineko	wuneko	wanike	wineke	winike	wineke	winike	puneke
horizontal	square	blue	hukeko	ponikio	huniko	huneko	huneko	hunike	kuneke	kuneke	huneke	ponike	hunekike
horizontal	square	red	pohumu	hukeko	ponekuko	poneko	puneko	punike	punike	puneke	puneke	ponike	punike
horizontal	triangle	black	muko	wakiki	kineko	wineki	wikeko	wineke	wineke	winike	wunike	winike	wineke
horizontal	triangle	blue	kokeguke	piniko	kuneko	kuneko	kuneko	hunike	kuneke	punike	honike	huneke	wineke
horizontal	triangle	red	kimu	koniki	pokiko	poneko	poneko	punike	punikiko	punike	punike	ponike	wineke
spiral	circle	black	kekewa	wiki	wiki	wikiko	winekiko	winikike	winikeke	winekike	winikike	winike	winekike
spiral	circle	blue	komuhuke	ponukiko	huki	kunekuki	kunkiko	kunikike	honekiko	honekike	kinike	ponike	hunikike
spiral	circle	red	kopo	ponikiko	poniki	poneko	pokiko	punikike	punekike	ponikike	poneike	ponike	punikiki
spiral	square	black	huwa	ponikiko	wineko	wikuki	winekiko	winekiko	winikike	winekike	winike	winikike	winikike
spiral	square	blue	hukike	hukeke	hunekiki	hunekiko	hunekiko	kunike	kunekike	kunekike	kinkike	hunike	punkike
spiral	square	red	ponikiko	ponikiko	ponekuki	poneki	puniko	punekiko	punikike	punekike	punkike	ponike	punikike
spiral	triangle	black	kowagu	winiko	wineki	winuki	wikiko	wanikike	winikike	winkike	winkeke	winike	winikike
spiral	triangle	blue	kokihuko	hukiki	hunekiko	hunekiko	kunekiko	kunekike	kinekike	kinekike	honkike	huneke	winike
spiral	triangle	red	kiwanike	kuniko	pokiko	ponekuki	pokiko	punikike	ponekiko	ponekiko	pinkike	punike	winikike

Table showing raw data for Chain D in Experiment III. Shaded cells indicate that this item was selected to be seen by the next generation during training.

Appendix B4

Raw data from 4 transmission chains in Experiment IV

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	kalu	kalu	lineki	lineki	huheki	huneki	huneki	huneti	kaneki	hineki	heniki
bounce	circle	blue	mola	balu	kineki	piteki	kiheki	kiheki	kineki	haneki	huneti	heniki	lineki
bounce	circle	red	pihukimo	capola	capola	kineki	haneki	haneki	haneki	kapeki	kineki	hiniki	hiniki
bounce	square	black	moki	lumoneki	mohuki	huteki	huteki	huteki	lineki	kaneki	kanitu	kaneki	haneki
bounce	square	blue	luneki	lineki	pinemahu	mahuki	kiteki	kiteki	haneki	kineti	linetu	kaneki	kaneki
bounce	square	red	lanepi	lanepi	huneki	hakeki	hateki	hateki	hapeki	hapeki	hanetu	kanetu	hineki
bounce	triangle	black	nane	kapola	kapiki	lineki	lineki	lineki	lineki	lineti	lineki	linitu	leneki
bounce	triangle	blue	kalakihu	mahiku	maheki	kineki	kineki	kineki	kiteki	kapeki	lineki	leniki	laniki
bounce	triangle	red	mokihuna	kapeki	mahetu	kapeki	hapeki	kapeki	kapeki	lineki	linetu	linitu	leniki
horizontal	circle	black	nelu	lumonamo	lumeno	kihetu	huniki	huniki	huniki	huniki	kiniki	heniki	heniki
horizontal	circle	blue	kanehu	humo	lumono	kakitu	kipiki	kihiki	kihiki	haniki	hiniki	hiniki	haniki
horizontal	circle	red	namopihu	lanehu	laneki	pitetu	haniki	haniki	haniki	kiniki	haniki	heniki	heniki
horizontal	square	black	lumonamo	lumonamo	kinehune	hutetu	hutiki	hutiki	haniki	kiniki	huniki	haneki	heneki
horizontal	square	blue	kinehune	lunepi	mahetu	hatetu	kiniki	kitiki	kitiki	kiniki	kaniki	kaneki	haneki
horizontal	square	red	lahupine	kinehune	kinehune	hatetu	hatetu	hatiki	hatiki	haneti	kaniki	kaneki	haneki
horizontal	triangle	black	kapihu	kapihu	kapetu	kapetu	liniki	liniki	liniki	liniki	liniki	liniki	liniki
horizontal	triangle	blue	humo	humo	kakitu	kakitu	kiniki	kiniki	kiniki	kapeki	kiniki	leniki	laniki
horizontal	triangle	red	lahupiki	capeki	pineku	katetu	hapiki	kapiki	katiki	kiniki	liniki	liniki	laneki
spiral	circle	black	pilu	pilu	pilu	pilu	hutetu	hunetu	hunetu	hunetu	hanetu	hinetu	hinitu
spiral	circle	blue	neki	neki	mahetu	pilu	kinetu	kihetu	kinetu	hapetu	hunetu	henitu	henitu
spiral	circle	red	pinemohu	pinemohu	pineku	pakiku	hanetu	hanetu	hanetu	hapetu	kinetu	hanetu	henitu
spiral	square	black	kilamo	neluki	hatuhi	hatuhi	hutetu	hutetu	linetu	kanetu	kanetu	kanetu	hanetu
spiral	square	blue	kahuki	kahuki	mahuki	lineku	kitetu	kitetu	kapetu	kinetu	hunetu	kanetu	hinetu
spiral	square	red	neluka	neluko	kineki	hatuhi	hatetu	hatetu	hapeku	kapetu	hanetu	hanetu	hanetu
spiral	triangle	black	luki	kahepi	kahepi	kapeti	huhetu	linetu	linetu	linetu	linetu	lenitu	linitu
spiral	triangle	blue	namola	luneki	luneki	kaketi	kinetu	kinetu	linetu	kinetu	linitu	linetu	lenitu
spiral	triangle	red	lumoka	neluki	kahetu	kapilu	hanetu	kapetu	kapetu	kinetu	linetu	lanitu	lanitu

Table showing raw data for Chain A in Experiment IV. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	melipa	wenumeko	mewenume	gahili	gahili	gahili	gahili	gahili	gahili	gahili	pagahilli
bounce	circle	blue	pame	pagahili	pagahili	pagahili	pagahili	pagahili	pagahili	pagahili	pagahili	pagahili	pagahilli
bounce	circle	red	mewega	mewemeko	mewemeko	gahili	megahili	megahili	megahili	megahili	megahili	megahili	megahilli
bounce	square	black	gamuwe	gamuho	mewenume	gahili	gahili	gahili	gahili	gahili	megahili	gahili	gahilli
bounce	square	blue	linuhiko	pawegame	pawegame	pagahili	pagahili	pagahili	pagahili	pagahili	pagahili	pagahili	pahilli
bounce	square	red	komehi	komehi	pawelinu	gahili	megahili	megahili	megahili	megahili	megahili	megahili	megahilli
bounce	triangle	black	hiko	wemenuko	mewehili	pahili	gahili	gahili	gahili	gahili	gahili	gahili	pagahilli
bounce	triangle	blue	palime	palime	palihili	pagahili	pagahili	pagahili	pagahili	pagahili	pagahili	pagahili	pagahilli
bounce	triangle	red	gawe	palime	paweganu	gahili	megahili	megahili	megahili	megahili	megahili	megahili	megahilli
horizontal	circle	black	hiwenuko	gamuho	menuko	galinu	linu	linu	linu	linu	linu	linu	wenu
horizontal	circle	blue	nuhiwenu	wemenuko	pagahili	pawenu	palinu	palinu	palinu	palinu	palinu	palinu	palinu
horizontal	circle	red	memenu	memenu	memenu	linu	melinu	melinu	melinu	melinu	melinu	melinu	wenu
horizontal	square	black	paweko	paweno	galinu	linu	linu	linu	linu	linu	linu	linu	galinu
horizontal	square	blue	konulipa	pawehili	pawega	linu	palinu	palinu	palinu	palinu	palinu	palinu	melinu
horizontal	square	red	linu	linu	linu	linu	melinu	melinu	melinu	melinu	melinu	melinu	malinu
horizontal	triangle	black	mume	pawehili	pali	linu	linu	linu	linu	linu	linu	linu	galinu
horizontal	triangle	blue	pawemeli	pawehili	lime	palinu	palinu	palinu	palinu	palinu	palinu	palinu	palinu
horizontal	triangle	red	liga	liga	palinu	linu	melinu	melinu	melinu	melinu	melinu	melinu	malinu
spiral	circle	black	melime	wegahili	mewenu	gawemu	wenu	wenu	wenu	wenu	wenu	wenu	wenu
spiral	circle	blue	munuko	himanuko	himanuko	pawenu	pawenu	pawenu	pawenu	pawenu	pawenu	pawenu	palinu
spiral	circle	red	komume	memenuko	memenuko	gawemu	gawemu	mewenu	mewenu	mewenu	mewenu	mewenu	mewenu
spiral	square	black	numekopa	mewuno	wemenu	wenu	wenu	wenu	wenu	wenu	wenu	wenu	wenu
spiral	square	blue	wega	wega	pawenu	mewenu	pawenu	pawenu	pawenu	pawenu	pawenu	pawenu	pamenu
spiral	square	red	higahili	hegahili	pawenu	mewenu	mewenu	mewenu	mewenu	mewenu	mewenu	mewenu	mewenu
spiral	triangle	black	pamu	memenuko	paweko	wenu	wenu	wenu	wenu	wenu	wenu	wenu	wenu
spiral	triangle	blue	gahi	paweko	paweko	pawenu	pawenu	pawenu	pawenu	pawenu	pawenu	pawenu	palinu
spiral	triangle	red	muwemeko	weganu	weganu	pawenu	mewenu	mewenu	mewenu	mewenu	mewenu	mewenu	mewenu

Table showing raw data for Chain B in Experiment IV. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	kuwo	lugana	luxana	lugana	peloana	logana	lagana	luxana	lugana	lugana	ligana
bounce	circle	blue	wonagi	pela	pela	pelo	luxana	hegana	lagana	hana	lugana	lugana	legena
bounce	circle	red	pelu	pela	luxana	napena	geki	nipana	lagana	h	lugina	lugena	lizana
bounce	square	black	wogipena	hika	hexigana	hexana	lugana	luxigana	luxana	luzana	luzana	luzana	ligana
bounce	square	blue	napena	napena	hexigana	peloana	peloana	luxana	luxana	luxana	luxana	luzana	luzana
bounce	square	red	penapiku	hika	luxana	napena	geki	geki	luxana	luxana	luzina	luzina	leguna
bounce	triangle	black	gapinahe	wociana	lugana	lugana	lugana	lagana	lugana	lugana	lugana	lugana	legana
bounce	triangle	blue	hewoku	luciwo	luciwo	luxana	luxana	luxana	lagana	lagana	lugena	lugina	legena
bounce	triangle	red	giku	gika	gika	geki	geki	geki	geki	geki	lugina	luzena	lizana
horizontal	circle	black	lugigipi	lugiana	hexigana	hexigana	hexigana	hena	hena	hena	hena	hena	nepana
horizontal	circle	blue	naheku	pela	hexigana	goana	hena	hena	hena	hena	huna	huna	huna
horizontal	circle	red	kuluwo	luxana	hena	hena	hena	hena	hena	hena	hina	hina	huna
horizontal	square	black	wogiluku	hexipena	hexipena	hexipena	hexigana	lagana	hexigena	hexigena	hena	nepana	nepana
horizontal	square	blue	gikuna	hika	hexigana	hipena	hexigana	hexigana	hexigana	hexigana	hexigana	nepena	nepena
horizontal	square	red	napeheku	gowo	hena	luxana	hexigana	hexigana	hexigina	hexigina	hexigina	hexigina	nezana
horizontal	triangle	black	penalu	penipika	higana	hipena	napena	lagana	nepena	nepena	hena	nepana	huna
horizontal	triangle	blue	pihena	hexigana	hexigana	napena	napena	napena	napena	nepena	nepana	nepana	huna
horizontal	triangle	red	naku	hena	hena	hena	napena	napena	nepena	hepena	nepina	nepina	huna
spiral	circle	black	lugana	lugana	luxana	lugana	hexipena	goana	goana	goana	pena	gouna	gouna
spiral	circle	blue	heku	peloana	peloana	pelo	goana	goana	goana	goana	guana	guana	guana
spiral	circle	red	wonalupe	wopelana	luxana	goana	goana	goana	goana	goana	goana	goana	goana
spiral	square	black	galukuna	hika	luxana	luxana	hexana	hexpina	hexigina	lugana	lagena	nepa	gouna
spiral	square	blue	napiwo	peloana	peloana	peloana	peloana	peloana	peloana	luxana	lagana	nepa	guzana
spiral	square	red	lupiwo	napena	napena	hexana	hexana	hexana	hexipena	lagana	lagana	gona	nepena
spiral	triangle	black	nahe	napena	lugana	hipena	hipena	hinepa	hipena	nepa	nepa	napa	napa
spiral	triangle	blue	pihe	goana	goana	luxana	heki	nepa	nepa	nepa	nupa	nepa	napa
spiral	triangle	red	galu	hika	hika	heki	heki	hinepa	nepa	nepa	napa	nepa	nepa

Table showing raw data for Chain C in Experiment IV. Shaded cells indicate that this item was selected to be seen by the next generation during training.

motion	shape	colour	0	1	2	3	4	5	6	7	8	9	10
bounce	circle	black	huhunigu	hokuhume	kikeko	wakeki	wakeki	wakeki	wakeki	wakike	wakiko	wakiko	wakiko
bounce	circle	blue	kemuniwa	hukeko	hukekuko	hukikuko	hunkeki	hunkeki	hukekuko	hukiko	hekiko	hekiko	pokiko
bounce	circle	red	kihupo	wagakiki	pokekuko	pokeki	pokeki	pokeki	pokeki	pokiko	pokiko	pokiko	pokiko
bounce	square	black	wakiki	wakiki	wakiki	wakiki	wakiki	wakiki	wakeki	wakeki	wakiko	wakiko	wakiko
bounce	square	blue	pokikehu	hokeko	kukeko	hukiki	hunkiki	hunkiki	hukekuko	hukeki	hekiko	hekiko	hekiko
bounce	square	red	waguhuki	nihu	ponikeko	pokikuko	pokiki	pokiki	pokekuko	pokeki	pokiko	pokiko	pokiko
bounce	triangle	black	nihu	nihu	kekuko	kekuko	wanikuko	wanikuko	wakiko	wakiko	wakiko	wakiko	wakiko
bounce	triangle	blue	niguki	niguki	nihu	hunikuko	hunikuko	hunikuko	hukikuko	hukiko	hekiko	hekiko	hekiko
bounce	triangle	red	koni	wagakiki	ponihumo	ponikuki	ponikuko	ponikuko	pokike	pokike	pokiko	ppokiko	pokiko
horizontal	circle	black	muwapo	wakeko	wakemo	wakemi	wakemo	wakemo	wakino	wakino	wanimo	wanimo	wanimo
horizontal	circle	blue	powa	hukoke	hukemo	hukemo	hunkemo	hunkemo	hukimo	hukimo	hekino	hekimo	henimo
horizontal	circle	red	hukinimu	koni	pokemo	pokemo	pokemo	pokimo	pokino	pokimo	ponimo	ponimo	ponimo
horizontal	square	black	wako	wako	wakemo	wakimo	wakimo	wakimo	wakino	wakimo	wanimo	wanimo	wanimo
horizontal	square	blue	hukeko	hukeko	hukimo	hukimo	hunkimo	hunkimo	wanimo	hukino	hekino	hekino	henimo
horizontal	square	red	pohumu	pohumo	pohumo	pokimo	pokimo	pokimo	pokino	pokino	ponimo	ponimo	ponimo
horizontal	triangle	black	muko	nihu	koni	wanimo	waknimo	wanimo	wanimo	wanimo	wanimo	wanimo	wanimo
horizontal	triangle	blue	kokeguke	wakiki	hunimo	hunimo	hunimo	hunimo	hunimo	hunimo	hekimo	hekimo	hekiko
horizontal	triangle	red	kimu	koni	pokoni	ponimo	ponimo	ponimo	ponimo	ponimo	ponimo	ponimo	ponimo
spiral	circle	black	kekewa	kekewa	wakeke	wakekuko	wakekuko	wakekuko	wakekuko	wakekuko	wahekiko	wakiheko	wahikeko
spiral	circle	blue	komuhuke	hokehume	hukeko	hukekoku	hunkekuko	hunkekuko	wakekuko	hukekuko	hekiheko	hekiheko	hehikeko
spiral	circle	red	kopo	ponikiko	pokikuko	pokikuko	pokekuko	pokekuko	pokekuko	pokekuko	pohekiko	pohiheko	pohikeko
spiral	square	black	huwa	kikuko	wakeke	wakikuko	wakikuko	wakikuko	wakikuko	wakekiko	wakiheko	wahiheko	wahikeko
spiral	square	blue	hukike	hokuhume	hukekuko	hukikuko	hunkikuko	hunkikuko	hukekuko	hukekiko	hekiheko	hekiheko	hehikeko
spiral	square	red	ponikiko	ponikiko	pokiko	pokikuko	pokikuko	pokikuko	pokekuko	pokekiko	pokiheko	pokiheko	pohikeko
spiral	triangle	black	kowagu	hokehume	niwakewa	wanikuko	wanikuki	wanikuki	wakikuke	wakikuko	wakiheko	wakiheko	wahikeko
spiral	triangle	blue	kokihuko	hukoke	hunikuko	hunikuko	hunikuki	hunikuki	hukikuko	hukikuko	hekiheko	hekiheko	hehikeko
spiral	triangle	red	kiwanike	ponikiko	pohumeko	pohikuko	ponikuki	ponikuki	pokikuko	ponikuko	pokeheko	pokiheko	pohikeko

Table showing raw data for Chain D in Experiment IV. Shaded cells indicate that this item was selected to be seen by the next generation during training.

Appendix B5

Raw data from 8 transmission chains in Experiment V

chain 1	0	1	2	3	4	5	6	7	8	9	10
	afa	afa	edd	efda	cda	cbfaf	edd	aff	ahf	fca	bfa
	bac	cee	cbfaf	efdb	bcfaf	bcfaf	dee	afaa	afaa	fhb	afc
	cde	bcfad	add	aee	edd	edd	aff	bchc	bchcf	afaa	ahc
	dcf	edd	efcba	bdd	cbfaf	dee	bcfb	bchcf	fhc	bchcf	ahf
	edd	bcdd	cbfa	add	add	add	bibf	bhbf	fah	ahf	fhc
	fcfd	fabee	cda	bcfaf	fdb	ebd	bcafa	bfcf	fbhf	bch	fca
	aefb	cbfaf	efda	cbfaf	efdb	afbf	ebd	cbf	bchc	afc	abff
	bbce	cbfa	bdd	cfaf	eca	cfbc	haf	chf	fch	ahc	fbhf
	caba	cbaee	bcfaf	bfaf	cfbd	afaa	fde	fhc	bch	bfa	bhcb
	deaf	cbd	aee	efaf	dee	aff	bchcf	hcf	aaf	bfc	bch
	edbcc	efad	efdb	cda	acfg	fdbc	hbc	haf	fca	fhc	bcbcf
	feadb	fdeba	bcfa	faf	cba	efbc	afaa	fha	afbf	ach	bchbf
	acbee	cbfad	cfaf	cdb	eda	haf	bcfcf	fah	hfc	abff	bchcf
	bdfef	ead	efaf	dee	efaf	bcfb	cafbf	hfbf	bcf	bhcb	afaa
	cbfad	cda	bfaf	cfba	cfaf	bibf	eac	fbhf	fbc	fbhf	fhb

Table showing raw data for Chain 1 in Experiment V.
The data is shown in the order in which it was produced by participants.

chain 2	0	1	2	3	4	5	6	7	8	9	10
	abd	bdc	aed	aaa	aaa	ccc	ccc	ccc	aaa	ccc	ccc
	bdc	ecd	fbfcf	ccc	fbfdf	dcf	aaa	aaa	ccc	dba	bbb
	cae	aabee	fcfbf	aca	fbfbd	aaa	dcf	fcdcf	fdbcf	fca	ddd
	def	fbfcf	aabee	aed	aca	fbfbd	fdbdf	facaf	fcdfa	aaa	aaa
	ecd	adba	ced	fcfbf	dfb	fdbdf	fdbfd	fdbfa	fcbdf	fcbaf	fcdaf
	fcdb	adb	ccfdb	fcfeb	dfbfd	fdacf	fcdfd	fcbdf	fdadf	fdbdf	fcbcf
	adba	ccbfc	aaa	bbdc	efb	dca	fdfdb	daf	fbdbf	fcbdf	fca
	bfde	ced	ccc	bed	cdf	fbdbf	fcdbf	caf	fdbaf	fcbcf	fdbdf
	ceaa	fcfbf	ccbdc	feb	cfbfd	fcdbf	dca	fadbf	fca	fcdbf	fcbaf
	dfed	ccbdc	aca	ccfbd	dfa	fdcdf	fbfbd	fdbdf	dba	fdcdf	cbd
	efdac	ccc	afa	ced	cac	dfbdf	fdafd	fdadf	fcadf	fabaf	dbfbd
	fbfcf	caf	abf	fdf	dfacf	fbd	dfbdf	fcadf	fbcaf	bbb	fcbab
	aabee	aed	bbdc	bbcd	dfc	fdfbf	fdcdf	fbdbf	fcacf	fcbfa	fcdcf
	bccfb	aaa	bcf	bedbb	ccc	fdfbd	daf	fca	fbcbf	fbdbf	dbdaf
	cebaf	afd	aec	fcfdf	fbddf	fcbdf	fdbcf	fdbcf	fcbaf	ddd	fcbad

Table showing raw data for Chain 2 in Experiment V.
The data is shown in the order in which it was produced by participants.

chain 3	0	1	2	3	4	5	6	7	8	9	10
	aec	bce	fdcdc	feg	bac	bac	fecac	bac	bac	eed	fecdad
	bce	fcfdf	eed	bac	fdcdc	fcg	ddde	eed	fecac	dde	eed
	cab	ddde	fecac	fdcac	fdede	fecac	dde	dde	dde	bac	dde
	dbd	fcfb	fec	fdcdc	fdcac	fedad	eeed	fecac	eed	fedad	fecda
	eff	eef	ddde	fecdc	facec	fecad	eed	fedad	fcedad	fedcd	fecae
	fefa	eca	fecdc	fdcec	fedcd	fedac	fdcdc	fecad	fceae	fedae	bac
	acbf	feca	bac	ddde	eeed	edd	fcdcd	feada	feeae	fcedad	fedad
	bfea	fecae	feg	eeed	ddde	dde	feadc	fedac	fecad	fed	fecad
	cead	eed	ebc	eed	dde	ddde	faedc	fde	fecda	fecdad	bfa
	ddde	dee	decdc	dde	eed	eeed	bac	fec	fec	fade	fceae
	ebafb	cgdf	fdcec	fcede	fcdad	fadad	fecad	fadcd	fedcd	fceda	fceda
	fcfdf	acbf	fcd	fcded	fcg	feaea	fedad	fdeae	fdaec	fecde	fcada
	abecd	bef	fdcac	fdece	fdced	fdcdc	fdaea	fcedad	feada	fce	fcede
	bacdc	fbcdc	cac	fcdcd	fecac	fcdcd	fec	dfeae	fedad	fcede	fcdad
	cbdae	bfd	dec	fecac	fcdac	fadec	fadcd	fadec	fcdad	fecae	fedae

Table showing raw data for Chain 3 in Experiment V.
The data is shown in the order in which it was produced by participants.

chain 4	0	1	2	3	4	5	6	7	8	9	10
	abc	fccf	adafb	bae	fceff	eaf	eaf	cfcea	baf	cdcef	baf
	baf	adafb	cba	eab	adacb	adacb	fcfeg	abf	fba	cdceg	bad
	cbd	bafb	dae	adafc	ebae	adacf	dadeg	cdcfa	efa	fba	cfde
	dce	fdeff	fdecb	fecff	eaf	fae	baf	ecbf	cdceg	baf	decf
	eea	fade	fdfcb	eade	eab	ebce	cfceg	adaef	dcdeg	cdef	cdcef
	fccf	fae	caeb	dae	adacf	fcfe	adaeb	dadeg	fcdb	fcde	cecdg
	abfa	eadcb	bae	cfe	ceaff	adaeg	fba	cfceg	fdce	def	dgef
	becd	cba	fdeb	adacb	acd	adaef	fcfda	fbce	cdcef	bad	cfeg
	cdfe	fdecb	bedfc	fdcff	fced	caf	fda	ecfb	cdef	decg	cdceg
	dfac	ecb	efb	bef	ceab	cea	cfcea	bfa	dcdef	decf	cgde
	ecbad	adec	edae	caeb	cfb	fcea	adaeg	eaf	cdcbf	cfde	decg
	fdeff	dafe	cfe	cfeb	ceaf	ebfe	adafc	fba	ebf	cdeg	cdge
	adafb	ade	cfeb	fced	fcea	fafeb	fceb	fcdb	cdfg	cecdg	gecd
	bedae	edbab	fdeff	cbae	bca	fafed	cfcda	fcb	dcda	cecdf	efcdg
	cebdb	efba	cdeff	feaff	fca	bae	aeb	caf	dadeg	fdce	cfedg

Table showing raw data for Chain 4 in Experiment V.
The data is shown in the order in which it was produced by participants.

chain 5	0	1	2	3	4	5	6	7	8	9	10
	daa	daa	daa	daa	bde	bde	fdfca	fdfac	fdfcb	fdfbc	add
	ecd	eaca	bde	bde	daa	daa	daa	daa	baa	cadbf	fdfbc
	feb	bed	ccafd	dece	dba	dba	dba	fdfbc	bda	fdfcb	fbdca
	afc	ccafd	edcd	bdfac	dbac	adfbd	dfdca	bda	bcfad	cadfb	fdfba
	bde	fbdef	dece	ccbdf	bdec	fdfba	bfcda	fdbca	bfdac	fbcad	fdfad
	cbaf	bde	edca	bdfca	dfdac	dfdca	afdcb	acdfb	bfdca	dbfca	fbdac
	dcec	dabf	fdebc	acfdb	afdcb	bdec	fafdc	fdfca	daa	bdfca	caa
	eaca	bdcf	bdfac	dfdbc	fdfac	dbc	bcdfa	dfb	adcfb	cbfda	cafbd
	fdbe	dcec	fbdfa	fbca	bdfac	adfcb	dfdac	bafdf	fdfbc	add	facbd
	abfb	cdbf	bdc	fdfca	ccbda	bdcfa	bdfac	bcfad	bdfac	caa	afcbd
	badfd	dafc	fbadc	deca	afdbc	bdce	cafda	cbfad	acdfb	daa	fcbad
	cceda	bad	dab	afdcb	bdcfa	bfdca	dbfac	bfdac	acdbf	cafbd	faa
	defac	badcf	dacbf	dba	bdcaf	afdcb	dafcb	cdfad	dfb	fbdca	cdd
	efbcb	cafbd	fdcbf	bdec	afcdb	dfdcb	cdafb	fdfcb	cadfb	fadcb	adbfc
	febdf	bacf	cbdaf	dbac	facbd	bdea	dfacb	baa	cadbf	fdfca	adbcf

Table showing raw data for Chain 5 in Experiment V.
The data is shown in the order in which it was produced by participants.

chain 6	0	1	2	3	4	5	6	7	8	9	10
	dba	ebfaf	dbda	dce	dce	ecd	ehc	ehc	ehc	daba	hdi
	ecd	cde	ebea	ebea	dbda	fdfb	ceh	fdabc	ehcaba	fdbc	dcd
	fab	fade	fbacb	abaea	ecd	daea	fdfb	edc	che	fdcd	fdcd
	aef	fbacb	cdea	edc	fcde	edbd	adaba	che	dadb	dhc	hdaba
	bdc	fcdf	cde	fced	dcda	fdbh	ecd	fbaec	fdcaba	cdeaba	faba
	cefb	cbade	edc	dbda	adaea	adab	fadba	edcba	fdcd	fdaba	fcdaba
	daed	dce	dce	ecd	ceg	ceh	edcd	echaba	daba	hdaba	fdaba
	ebea	fce	cbae	cfebc	fdfb	ehc	fdaba	fdaba	fdaba	fdec	ecd
	fcde	fcdce	abaea	eaeb	bedc	abeda	fbabh	dbda	edc	dcaba	ech
	affc	acede	ecd	fcea	ebea	dbda	aeabc	adab	edcaba	adab	dcaba
	bcbda	ebea	cabec	dcea	fcfab	aeaba	aeaba	adbc	fdbc	fcdaba	fabdcd
	caddc	fad	dcea	fecba	edea	edaba	fdaeb	abafd	bdch	hdi	ceaba
	dfece	adc	eaeb	ced	cde	cded	aced	abach	abad	eab	fhaba
	ebfaf	dbda	fcad	bcefb	efc	dcea	ecdad	fbda	fbch	feaba	fedh
	fbacb	acde	fced	feabe	aeaba	adabe	dbda	fbdad	edaba	fheaba	adcd

Table showing raw data for Chain 6 in Experiment V.
The data is shown in the order in which it was produced by participants.

chain 7	0	1	2	3	4	5	6	7	8	9	10
	dfb	bfacc	caedb	dga	faff	dgc	ghc	gdc	bfbb	bfb	aea
	eea	ceadb	adb	faff	afaa	faff	ged	ged	babb	abab	bfb
	fcd	feb	faee	fabb	dgb	afaa	gde	babb	fbf	cefgd	cefgd
	abc	bad	cadea	caedb	egc	egd	fbff	bfb	fecgd	cefdg	cefdg
	bde	fde	facc	cadeb	facbe	facdg	faff	bfbb	bba	ceb	cde
	cefb	cabde	feb	cabdc	ebd	bfaa	bfaa	bfaa	bfb	ced	baba
	dbef	bbd	bda	ebd	gde	edg	gdc	ced	bab	bbf	dcdeg
	ecff	afee	fabb	edb	fabb	fbff	fcedg	cedfg	cedfg	bfbf	bfbf
	fcba	eff	acde	bec	bfaa	ghc	bfbb	abaa	ced	aea	fde
	afee	fedc	aff	dec	ceg	ged	babb	gcegd	ceb	dgedc	bab
	bfacc	fadb	cadec	afaa	aedgc	bfd	abaa	cdged	cadb	beb	faf
	caeda	adb	fad	afbb	dgc	abaa	afaa	babf	aedb	cef	cfe
	dacbd	fbdd	dbe	afbc	egK	bea	faf	bfbf	aafb	baba	cdfgd
	ebadc	caeda	faff	cabec	ceba	gdc	bfb	gefdg	ffab	fea	cgdfg
	faddb	efb	dga	fac	babb	abb	fbf	fbf	baba	cde	cda

Table showing raw data for Chain 7 in Experiment V.
The data is shown in the order in which it was produced by participants.

chain 8	0	1	2	3	4	5	6	7	8	9	10
	bad	efdf	cbc	caccf	cfcca	dfdfa	bae	dfdfa	dfdfa	abe	abe
	cbc	bfdfa	cac	abe	ebfb	ebfb	bfe	fbe	dfabc	fbe	fbe
	eda	eccdf	dad	bcea	fbe	dfdca	bfa	afb	bfc	dfdfc	dcdcf
	afb	cbc	fdcfa	fbed	eaf	ebf	dcdca	bfa	baebf	dfabe	dabcf
	dbe	cac	bead	efdfc	eab	dcdca	dfdfa	bae	dcdcf	bfc	dcabe
	febd	caccf	abe	becd	dfdfa	efdca	cacfb	dcdca	abf	dcdcf	dafbe
	bbab	dad	caccf	dfdfa	dfdcf	fbe	fbe	dcdfa	afabe	abc	bfc
	cecf	eddfa	dfefa	fbea	dfeb	cfcca	efd	fab	dcabf	bcf	dadcf
	efdf	dea	dfcfa	fcafc	edfb	bfa	ebd	cab	fba	dcdbf	dcabf
	acdd	dfabe	fbea	befb	efdcf	efca	fba	dadfa	abe	dcbcf	dcbcf
	deabe	bead	fdfca	fecfa	efdca	efbf	afb	baebf	fbe	dcabe	dcf
	fafcc	fcbf	edcdf	dfdcf	ecfe	bae	efdab	bca	dfdaf	dfabc	dfdcb
	bfdfa	efdfa	bced	efcfd	edca	ecfca	cfdca	dfabc	dfb	fbc	dcfbe
	caeea	fded	fbec	edfb	efba	efcfa	efbfd	acdfb	bae	dabfc	dcace
	ecdcf	bced	dfec	ebfb	dcdcf	fbae	cfdea	fda	fab	dcfbe	dbdac

Table showing raw data for Chain 8 in Experiment V.
The data is shown in the order in which it was produced by participants.

Appendix C

Instructions to participants in Experiment V

Thank you for agreeing to participate in this study.

During the experiment you will see a series of letter strings appear on the screen. We would like to see how well you can learn them.

After each string appears, there will be a short delay before you are allowed to type in what you think you saw. Try to remember the strings as accurately as possible. You can use the backspace button if you make a mistake, and can press ENTER to see the next string.

Please press the track-pad button when you are ready to begin, and good luck.

Thank you! You just saw 15 different strings.

We would like you to try to recall all of them now as best you can.

Please keep going until you have tried to remember each one. We will give you an indication of how many you have left to enter. However, we won't tell you how many you got right until the end of the experiment.

Press the track-pad button to begin.

You have already entered this string.
Please try again.

Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language

Simon Kirby*[†], Hannah Cornish*, and Kenny Smith[‡]

*School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh EH8 9LL, United Kingdom; and [‡]Division of Psychology, Northumbria University, Newcastle-upon-Tyne NE1 8ST, United Kingdom

Edited by Dale Purves, Duke University Medical Center, Durham, NC, and approved June 6, 2008 (received for review August 20, 2007)

We introduce an experimental paradigm for studying the cumulative cultural evolution of language. In doing so we provide the first experimental validation for the idea that cultural transmission can lead to the appearance of design without a designer. Our experiments involve the iterated learning of artificial languages by human participants. We show that languages transmitted culturally evolve in such a way as to maximize their own transmissibility: over time, the languages in our experiments become easier to learn and increasingly structured. Furthermore, this structure emerges purely as a consequence of the transmission of language over generations, without any intentional design on the part of individual language learners. Previous computational and mathematical models suggest that iterated learning provides an explanation for the structure of human language and link particular aspects of linguistic structure with particular constraints acting on language during its transmission. The experimental work presented here shows that the predictions of these models, and models of cultural evolution more generally, can be tested in the laboratory.

cultural transmission | iterated learning | language evolution

The emergence of human language has been cited by Maynard Smith and Szathmary (1) as the most recent of a small number of highly significant evolutionary transitions in the history of life on earth. The reason they give for including language in this list is that language enables an entirely new system for information transmission: human culture. Language is unique in being a system that supports unlimited heredity of cultural information, allowing our species to develop a unique kind of open-ended adaptability.

Although this feature of language as a carrier of cultural information obviously is important, we have argued that there is a second sense in which language is an evolutionary milestone: each utterance has a dual purpose, carrying semantic content but also conveying information about its own construction (2–5). Upon hearing a sentence, a language learner uses the structure of that sentence to make new inferences about the language that produced it. This process allows learners to reverse-engineer the language of their speech community from the utterances they hear. Language thus is both a conveyer of cultural information (in Maynard Smith and Szathmary's sense) and is itself culturally transmitted. This cultural transmission makes language an evolutionary system in its own right (2–3), suggesting another approach to the explanation of linguistic structure. Crucially, language also represents an excellent test domain for theories of cultural evolution in general, because the acquisition and processing of language are relatively well understood, and because language has an interesting, nontrivial, but well documented structure.[§]

During the past 10 years a wide range of computational and mathematical models have looked at a particular kind of cultural evolution termed "iterated learning" (4–13).

Iterated Learning. Iterated learning is a process in which an individual acquires a behavior by observing a similar behavior in another individual who acquired it in the same way.

Spoken (or signed) language is an outcome of iterated learning. Although in some circumstances aspects of language may be explicitly taught, acquired from a written form, or arise from deliberate invention, almost all the features of the languages we speak are the result of iterated learning. Models of this process (4–13) demonstrate that, over repeated episodes of transmission, behaviors transmitted by iterated learning tend to become 1) easier to learn, and 2) increasingly structured. Note that this process is cumulative and is not considered to arise from the explicit intentions of the individuals involved. Rather, this type of cultural evolution is an "invisible hand" process leading to phenomena that are the result of human action but are not intentional artifacts (14).

Although these models are indicative of the power of cultural evolution in explaining language structure, skepticism remains as to how well computational models of learning match the abilities and biases of real human learners. For example, responding to a growing body of computational models of the emergence of multiword utterances from unstructured randomness (5, 8, 10, 11, 15), Bickerton notes, "Powerful and potentially interesting although this approach is, its failure to incorporate more realistic conditions (perhaps because these would be more difficult to simulate) sharply reduces any contribution it might make toward unraveling language evolution. So far, it is a classic case of looking for your car-keys where the street-lamps are" (16, p. 522).

What is needed, therefore, is an experimental paradigm for studying the evolution of complex cultural adaptations using real human participants. Ideally, this paradigm should mirror previous computational and mathematical models and provide a test for the claim that iterated learning leads to adaptively structured languages. It should demonstrate whether cumulative adaptive evolution without intention is possible purely by virtue of cultural transmission.

In this paper, we implement such a paradigm and demonstrate cumulative, adaptive, nonintentional cultural evolution of an artificial language in a laboratory population of human participants.

Diffusion Chains. Diffusion-chain studies provide the best example of experimental treatments of iterated learning. In these experiments a participant observes some target behavior (provided by the experimenter) and then is required to replicate that behavior in some way that can be observed by a second participant. This second participant in turn attempts to replicate the first participant's

Author contributions: S.K., H.C., and K.S. designed research; H.C. performed research; S.K., H.C., and K.S. analyzed data; and S.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[†]To whom correspondence should be addressed. E-mail: simon@ling.ed.ac.uk.

[§]From a practical perspective it is also an ideal subject for study in that it is relatively straightforward to record and analyze precisely.

This article contains supporting information online at www.pnas.org/cgi/content/full/0707835105/DCSupplemental

behavior for a third participant, and so on. (We refer to each iteration of this cycle as "1 generation.") Using this procedure, we can observe the diffusion of behavior through a chain of cultural transmission. The first reported use of this methodology was by Bartlett in 1932 (17), but only recently did researchers begin to apply this approach systematically (18–24)

The most recent, and arguably the most significant, instance of a diffusion-chain experiment is the work of Horner *et al.*, which explores the cultural transmission of tool-use strategies in populations of chimpanzees and children (24). Diffusion chains are set up in which an experimenter demonstrates 1 of 2 possible techniques for opening a puzzle box ("artificial fruit") to a participant. Subsequent participants observe their predecessor's box-opening behavior and then in turn become the model for the next generation. These experiments demonstrate clearly that both chimpanzees and children are capable of high-fidelity cultural transmission: the box-opening technique used by the last participant in the chains (of up to 10 individuals) is the same as that demonstrated to the first participant, with a chain of faithful transmission between the first and last participants.

Although these experiments show that cultural transmission can be studied empirically even in nonhumans, they do not support our claim that culture leads to cumulative nonintentional adaptation because the behavioral information that is being transmitted is drawn from a limited set of possibilities. For example, in the puzzle-box study, there are essentially 2 different strategies for opening the box. The task is not complex enough to demonstrate adaptation, let alone cumulative adaptation. In any case, both the strategies seem to be equivalently "adaptive" in cultural and environmental terms, in that both open the box and both are transmittable.

To get around these problems and to allow us to make a direct comparison with human language, we replicate the basic diffusion-chain design with a more complex artificial-language learning task of labeling visual stimuli with strings of written syllables (25, 26). To make this task tractable, we use adult human participants and observe the cultural evolution of the artificial language for 10 cultural generations.

This work bears some resemblance to a recent body of experimental work on the shared construction of communication systems (27–30). Of particular relevance is a recent paper by Selten and Warglien (30) that demonstrates that pairs of participants sometimes can create structured and efficient communication systems over the course of repeated interactions. The major difference between the experiments described here and the work of Selten and Warglien is the role of intentional design. In Selten and Warglien's experiments, as in those of Galantucci (27) and Garrod *et al.* (28, 29), participants interact repeatedly with the explicit goal of arriving at a shared system for communication. Therefore the systems they construct are the outcome of conscious design. Our diffusion-chain experiment allows us to explore whether structured languages can emerge without intentional design, as has been argued to be the case for language (14).

Design of Experiment 1. Participants are asked to learn an "alien" language made up of written labels for visual stimuli. The stimuli are pictures of colored objects in motion, and the labels are sequences of lowercase letters (see Fig. 1 for an example and the *Methods* section for more details).

For training purposes, the language to be learned (a set of string–picture pairs) is divided randomly into 2 sets of approximately equal size: the SEEN set and the UNSEEN set. A participant is trained on the SEEN set, being presented repeatedly with each string–picture pair in random order (see *Methods* for details). During subsequent testing, participants are presented with a picture and asked to produce the string they think the alien would give for that picture. Participants are tested on both the SEEN and UNSEEN sets in their entirety.

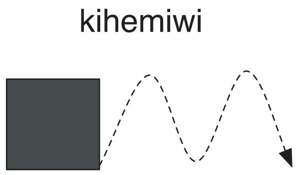


Fig. 1. An example string–picture pair.

The initial set of labels in the language is generated and assigned randomly, and the first participant in the experiment is trained on this random language. Subsequent participants are trained on the output of the final testing of the previous participant, which is re-divided into new SEEN and UNSEEN sets. Note that the experimental procedure is equivalent for all participants, despite the different sources of training data: at no stage are participants told that they are being trained on the output of another person, nor did any participants guess that the transmission of an acquired language was part of the experiment. Crucially, participants believe they are copying the input language as best they can; a posttest questionnaire revealed that many participants did not even realize that they were being tested on stimuli they had not seen in training, so that intentional design on the part of the participants is unlikely. To put it another way, the participants' goal is to reproduce the language, not improve to it in some way. (We return to this point in the *Discussion* section).

Our hypothesis is that we will observe cumulative adaptive evolution of the language being transmitted in this experiment; that is, we should see the emergence of adaptive structure in response to the pressure on the language to be transmitted faithfully from generation to generation. If this hypothesis is correct, we should see 2 things: 1) an increase in the learnability of the language over generations (i.e., a decrease in transmission error), and 2) the evolution of linguistic structure (i.e., an increase in predictability in the mapping between meanings and signals).

We devised 2 measures to test this hypothesis. First, we used a measure of string similarity to compare words in the languages of participants at adjacent generations (see *Methods*). The Levenshtein edit distance (31) between pairs of words (i.e., the smallest number of character insertions, replacements, and deletions required to transform 1 word into the other) provides a reasonable theory-neutral measure of distance. We normalized the edit distance for length of words so that identical strings have a distance of 0 and maximally distinct ones have a distance of 1. The mean distance between all of the words in a participant's output and the corresponding words in the previous generation's output gives a straightforward measure of the error in transmission of the language.

Second, we constructed a measure of linguistic structure based on measures of compositionality used in some computational models (12). Our aim was to quantify the degree to which the mapping between meanings (visual scenes) and signals (character strings) is systematic, an obvious hallmark of structure in human language. A language is systematic if patterns of similarity and dissimilarity in signals provide information about the relationship between the meanings those signals map on to. Accordingly, we calculated the correlation between all pairs of edit-distances in the set of signals and the corresponding distances between meanings (i.e., whether they differed in shape, color, and/or movement). By using Monte-Carlo techniques, we can calculate the extent to which this alignment between meaning and signal differs from the alignment we would expect to see by a random, unstructured assignment of signals to meanings (see *Methods* for details).

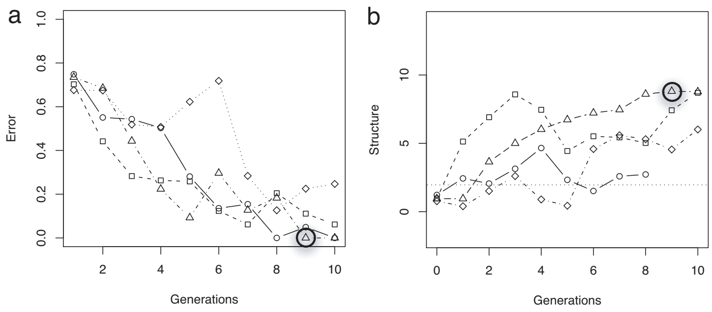


Fig. 2. Transmission error and a measure of structure by generation in 4 chains *a* shows the increase in learnability (decrease in error) of languages over time *b* shows structure in the languages increasing. The dotted line in *b* gives the 95% confidence interval so that any result above this line demonstrates that there is a nonrandom alignment of signals and meanings. In other words, structure in the set of signals reflects structure in the set of meanings. In 2 cases, this measure is not defined and therefore is not plotted (see *Methods*). The language discussed in the paper is circled.

Results of Experiment 1. The results of our first experiment, involving 4 separate diffusion chains of 10 participants each, are shown in Fig. 2. Each of these chains was initialized with a different random language. There is a clear and statistically significant decrease in transmission error between the initial and final generations (mean decrease 0.748, SD = 0.147; $t(3) = 8.656$; $P < 0.002$). This decrease confirms the first of our predictions: the language is adapting to become increasingly transmissible from generation to generation. Indeed, toward the end of some chains the language is transmitted perfectly: these participants produced exactly the same strings for every meaning as their predecessor, although they had not been exposed to the strings associated with half of those meanings.

How is this adaptation possible? Is any structural evolution of the language taking place as in the second of our 2 predictions? As Table 1 shows, the number of distinct strings in each language decreases rapidly. The initial random languages are completely unambiguous: every meaning is expressed by a distinct signal. The transmission process cumulatively introduces ambiguity as single strings are re-used to express more and more meanings. In other words, the languages gradually introduce underspecification of meanings. Clearly, the reduction in the number of strings must make a language easier for participants to learn, but the reduction alone cannot account for the results we see. For example, the reduction does not explain how, in some chains, participants are able to produce the correct signal for every meaning, including meanings drawn from the UNSEEN set.

The answer to this puzzle lies in the structure of the languages. The initial random language is, by definition, unstructured: nothing in the set of signals gives any systematic clue to the meanings being conveyed. The only way to learn this language is by rote. Equally, if a language is randomly underspecified, then rote learning is the only way it can be acquired. For example, if the same signal is used for a black spiraling triangle and a red bouncing square, then a learner must see this signal used for both of these meanings to learn it. Because we deliberately hold items back from the SEEN set, rote learning for all meanings is impossible. For learners to be able to generalize to unseen meanings successfully, there must be systematic underspecification.

We can observe exactly this kind of structure evolving by examining a language as it develops in the experiment. For example, by generation 4 in 1 of the diffusion chains, the string *tuge* is used exclusively for all pictures with an object moving horizontally. The distribution of the other strings in the language is more idiosyncratic and unpredictable at this stage. By generation 6, *poi* is used to refer to most spiraling pictures, but there are exceptions for triangles and squares. Blue spiraling triangles or squares are referred to as *tupin*, and red spiraling triangles or squares are *tupim*. In the following generation, these exceptional cases are reduced to the blue spiraling triangle and the red spiraling square. By generation 8 (shown in Fig. 3), and also for generations 9 and 10, the language has settled on a simple system of regularities whereby everything that moves horizontally is *tuge*, all spiraling objects are *poi*, and bouncing objects are divided according to shape.

It is precisely because the language can be described by using this simple set of generalizations that participants are able to label correctly pictures that they have never previously seen. This generalization directly ensures the stable cultural transmission of the language from generation to generation, even though each learner of the language is exposed to incomplete training data.

Table 1. Number of distinct words by generation in the first experiment

Generation	0	1	2	3	4	5	6	7	8	9	10
○ Chain 1	27	17	9	6	5	4	4	2	2	2	2
□ Chain 2	27	17	15	8	7	6	6	6	5	5	4
△ Chain 3	27	24	8	6	6	5	6	5	5	5	5
◇ Chain 4	27	23	9	10	9	11	7	5	5	4	4

Symbols correspond to those in Fig. 2.

Fig. 3. An example evolved language in the first experiment. This language exhibits systematic underspecification, enabling learners to reproduce the whole language from a fragment.

Table 2. Number of distinct words by generation in the second experiment

Generation	0	1	2	3	4	5	6	7	8	9	10
○ Chain 1	27	23	22	17	21	21	17	21	25	13	16
□ Chain 2	27	26	13	10	10	16	16	12	12	13	12
△ Chain 3	27	11	16	14	12	17	14	16	20	19	12
◇ Chain 4	27	19	19	17	19	17	22	23	21	27	23

Symbols correspond to those in Fig. 4.

Our structure measure confirms that the languages evolve to become more structured. As can be seen in Fig. 2*b*, significantly nonrandom structure in the mapping from meanings to signals emerges rapidly. Furthermore, the languages produced by the final generation are significantly more structured than the initial languages (mean increase 5.578, SD = 2.968, t (3) = 3.7575, $P < 0.02$).

Languages in this experiment are evolving to be learnable, and they are doing so by becoming structured. This development of structure confirms our hypothesis regarding the cultural evolution of language. However, we are interested in whether it would be possible for a language to evolve that is learnable and structured but also expressive, i.e., a language that would be able to label meanings unambiguously. Such a language cannot rely on systematic underspecification of meanings but instead must find some other means of gaining structure.

Design of Experiment 2. Accordingly, in the second experiment we made a single minor modification: we "filtered" the SEEN set before each participant's training. If any strings were assigned to more than 1 meaning, all but 1 of those meanings (chosen at random) was removed from the training data. This filtering effectively removes the possibility of the language adapting to be learnable by introducing underspecification: filtering ensures that underspecification is an evolutionary dead-end. This process, although artificial, is an analogue of a pressure to be expressive that would come from communicative need in the case of real language transmission.

Results of Experiment 2. As expected, under the modified regimen, the overall number of words in participants' output remains comparatively high throughout the experiment, as shown in Table 2. Fig. 4*a* shows how transmission error changes as the language evolves. Once again, it is clear that the languages are becoming more learnable over time (mean decrease 0.427, SD = 0.106, t (3) = 8.0557, $P < 0.002$) although it is not possible to introduce the kind

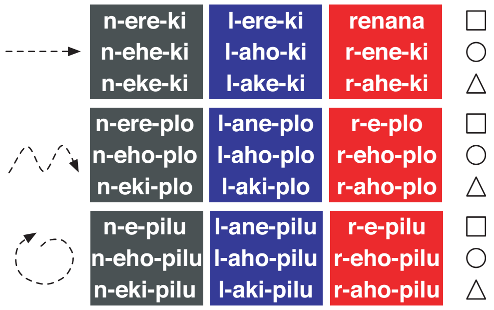


Fig. 5. An example evolved language in the second experiment. The language is structured: the string associated with a picture consists of substrings expressing color, shape, and motion, respectively. The hyphens represent 1 way of analyzing the substructure of these strings and are added purely for clarity; participants in the experiment always produced strings of characters without spaces or any other means of indicating substructure.

of underspecification seen in Experiment 1. Furthermore, it is clear from Fig. 4*b* that, as in Experiment 1, the languages are becoming increasingly structured over time (mean increase, 6.805, SD = 5.390, t (3) = 2.525, $P < 0.05$). Because filtering rules out the generalizations that emerged in the previous experiment, a different kind of structure that does not rely on underspecification must be emerging.

If we examine the languages at particular stages in their cultural evolution, we can see exactly what this structure is. For example, Fig. 5 shows the language output by a participant at generation 9 in 1 of the diffusion chains. When one looks at this language, it immediately becomes clear that there is structure within the signals. We can analyze each signal as 3 morphemes expressing color, shape, and movement, respectively, with 1 exceptional irregularity (*renana* for a bouncing red circle). It turns out that this general structure emerges by at least generation 6 and persists to the end of the experiment, although the details change as some morphemes are lost or are reanalyzed from generation to generation [see supporting information (SI) Tables S1–S8 for the complete set of languages].

Discussion

What we have observed here under laboratory conditions is cumulative cultural adaptation without intentional design. Just as

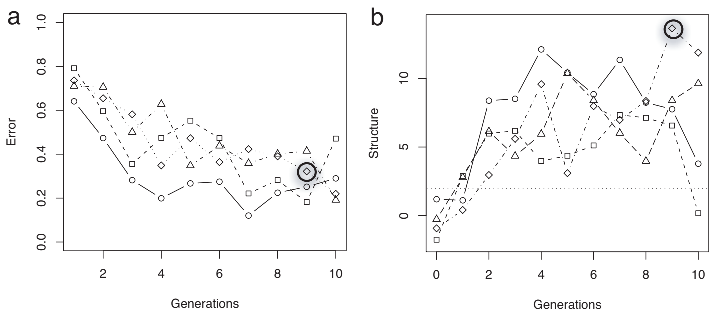


Fig. 4. Transmission error and structure by generation in the experiment in which ambiguous data were removed from the training set at each generation. *a* gives error for the whole language; *b* gives structure. These results show that, despite the blocking of underspecification, structure still evolves that enables the languages to become increasingly learnable. The language discussed in the paper is circled.

previous computational models have predicted (4–13), the culturally evolving language has adapted in a way that ensures its successful transmission from generation to generation, despite the existence of a bottleneck on transmission imposed by the incomplete exposure of each participant to the language. Cultural adaptation results in languages that circumvent this transmission problem by exploiting structure in the set of meanings to be conveyed. Note that this adaptation is cumulative with respect to learnability and structure but not with respect to expressivity: cumulative adaptation does not suggest that the languages necessarily become more functional with respect to communication.

In all our experiments we have shown that languages, by virtue of being culturally transmitted, become increasingly learnable and increasingly structured. An obvious question is: to what extent does the structure we see emerging resemble structures found in real human languages?

In the first experiment, we saw underspecification introduced into the language. This underspecification was not random but was systematic, in that similar meanings were given the same label. The form of the language reflected regularities in the visual scenes, namely that they consisted of shape, color, and motion. Of course, in the experiment this process ran unchecked and in some cases led to languages in which almost every meaning was expressed by a single signal.

The languages in our first experiment therefore could be seen as being counter-functionally ambiguous. However, there is another way of thinking about our results. Rather than seeing the emerging language as ambiguous, some participants thought it revealed something about the way the aliens saw the world. For example, in posttest discussions, 1 participant noted that "color is not important to these aliens." This observation suggests that the participants did not consider the language to be ambiguous, but instead thought that it reflected the distinctions in meaning that the aliens were interested in communicating. The collapse of distinctions based on color (which eventually occurred in all 4 replications of the first experiment) in favor of distinctions based on shape and movement is compatible with the literature on a shape bias, an expectation that words will refer to shapes of objects rather than to properties such as color or texture (32). It may be that, while adapting to become more learnable by eliminating semantic distinctions, the languages in the experiment retain the distinctions that seem most salient and/or likely to be labeled linguistically.

Systematic underspecification similar to that found in the experiments is an important feature of natural language. For example, in the class of nouns only proper names refer to specific entities. Other nouns are underspecified and typically correspond to natural classes. However, systematic underspecification is not the only way in which the structure of the set of meanings makes itself felt in linguistic expressions. Most obviously, natural languages exhibit the species-unique property of compositionality in syntax and morphology.[¶] The meaning of an expression normally is a function of the meanings of subparts of that expression and of the way the subparts are put together. It is precisely this property that we hypothesize allows language to be both learnable and expressive.

Expressivity in human language is assumed to be a consequence of the use of language for communication and also may be attributable to predispositions of child language learners (33, 34). In 1 computational model of iterated learning (8), an expressivity requirement is enforced simply by filtering out ambiguous meaning-strings from the data given to the learner, leaving a training set with a unique 1-to-1 mapping between meanings and strings. Although learners still are free to infer ambiguous strings, such ambiguity would not be transmitted to the following generation.

¶Arguably, the dance of honey bees (35) and the calls of Campbell's monkeys (36) are both minimally compositional. However, there is no evidence (as yet) for culturally transmitted or open-ended compositional communication outside our species.

We implemented exactly this filtering process in the second experiment, to dramatic effect, even though for the participants the conditions in this experiment were essentially identical to those in the previous experiment. As in Experiment 1, after being presented with string–picture pairs, the participants had to recall these pairs and generalize to unseen pictures. Nevertheless, unlike in the previous experiment, systematic compositional structure emerged. Rules evolved for constructing signals out of a combination of meaningful substrings, and these rules tended to be transmitted from generation to generation once they had emerged (seeTables S1–S8 for the full set of languages). The difference between these 2 experimental settings is simply that the second introduces a new adaptive challenge for the evolving language. To be transmitted faithfully from generation to generation, a language in this experiment must be both learnable and unambiguous. The learnability constraint is imposed by the participants in the experiment, and the ambiguity constraint is imposed by our additional filter.

The result is the evolution of exactly the type of structure that optimizes both these competing constraints: compositionality. The evolution of this structure reveals a key feature of cultural transmission: it gives rise to adaptive systems that respond to the pressures imposed by the transmission bottleneck that exists between the producer and learner of behavior. Crucially, this adaptation by the language maximizes its own transmissibility, and the adaptation can take place without intentional design on the part of the individuals involved. Participants in the second experiment could not be aware that ambiguous signals were being filtered, and yet a completely different sort of structure emerged. This finding demonstrates that adaptation can be independent of the intentions of individuals.

Finally, the difference between the 2 experiments also shows that the languages that emerge are not simply a reflection of the native language of the participants. A participant's first language may influence the learnability of a particular artificial language and therefore play a role in shaping the cultural evolution of those languages in our experiments. However, this explanation cannot be the whole story: if participants were merely stamping their own linguistic knowledge onto the data that they were seeing, there would be no reason we would find rampant structured underspecification in the first experiment and a system of morphological concatenation in the second.

Conclusions

We have shown that it is possible to study cumulative cultural adaptation in the laboratory. Using a diffusion-chain paradigm with an artificial-language learning task, we provide empirical support for computational and mathematical models of iterated learning that show language to be an adaptive system in its own right. We demonstrate the cumulative evolution of an adaptive structure without intentional design on the part of the participants in the experiment.

We can understand the linguistic structure emerging in these experiments as an adaptive response by language to the problem of being transmitted from generation to generation. In particular, language faces the problem of being reproducible from a subsample. In the first experiment, the language solves this problem by introducing systematic underspecification in the meaning-signal mapping. In the second experiment, the language faces the additional challenge of being transmitted despite filtering for ambiguity. Compositional structure is a potential solution to this particular transmission problem, and this structure emerges. It is important to reiterate that participants in the experiment did not intentionally design this solution; indeed, they were not even aware of the problem. Participants believed they were reproducing as best they could the language to which they were exposed. Just as biological evolution can deliver the appearance of design without the existence of a designer, so too can cultural evolution.

Methods

Eighty participants were recruited to participate in an ''alien language'' learning study. Each had to learn a language made up of written labels for visual stimuli. Participants were university students with no background in linguistics. The female:male ratio was 46:34, the mean age was 22.5 years, the minimum age was 18 years, and the maximum age was 40 years. The experiment was conducted in accordance with the ethics procedures of the Department of Linguistics and English Language at the University of Edinburgh. Participants carried out the experiment at a computer terminal and received written and verbal instructions (see *SI Text*). During training, participants were presented with string–picture pairs on the computer monitor. During testing, participants were presented with pictures on the monitor and were prompted to enter strings using the keyboard, with any sequence of alphanumeric characters being permissible.

Visual Stimuli. There were 27 possible stimuli to be labeled. Each was a colored object with an arrow indicating motion. Each object feature (shape, color, motion) varied over 3 possible values: square, circle, or triangle; black, blue, or red; horizontal motion, bouncing, or spiraling motion.

Labels. The set of labels in the initial language was generated and assigned randomly and was constructed by concatenating between 2 and 4 syllables (without spaces between) taken from a set of 9 simple consonant–vowel pairs. Because participants were free to enter any sequence of characters they chose during testing, subsequent labels were unconstrained.

Training and Testing Regimen. Each language (a set of 27 string–picture pairs, 1 string for each of 27 possible pictures) was divided randomly into 2 sets: the SEEN set (14 string–picture pairs) and the UNSEEN set (13 string–picture pairs). Each participant acquired the language in a single session comprising of 3 rounds of training with an optional 2-minute break between rounds. A single round of training consisted of 2 randomized exposures to the SEEN set, followed by a test. In the first 2 rounds this test phase contained only half the SEEN and half the UNSEEN items; the final test at the end of the third round (which was the only source for the next generation's language) consisted of all 27 pictures.

During each training pass through the SEEN set, participants were presented with each pair in a random order, with the string being displayed for 1 second followed by both string and picture being displayed for a further 5 seconds. During testing, participants were presented with a picture and prompted to type in the string they thought the alien would produce for that picture.

In the second experiment, the SEEN set was filtered before presentation to participants. Specifically, if any string labeled more than 1 picture, all but 1 of those string–picture pairs (chosen at random) was moved into the UNSEEN set. As a result, the training data seen by participants in the second experiment consisted of a purely 1-to-1 mapping from strings to pictures, even if the language of the previous generation included 1-to-many mappings.

Diffusion-Chain Design. The first participant in the experiment was trained on a language with randomly constructed labels. Subsequent participants were trained on the output of the final testing of the previous participant: the previous participant's final testing output was randomly redivided into a new SEEN and UNSEEN set.

Measure of Transmission Error. The mean distance between all the signals in a participant's output and the corresponding signals in the previous generation's output gives a measure of intergeneration transmission error, and is given by

$$E(i) = \frac{1}{|M|}\sum_{m \in M} \mathrm{LD}(s_i^m, s_{i-1}^m)$$

where s_i^m is the string associated with meaning m by the participant at generation i, LD s_i^m, s_j^m is the normalized Levenshtein distance (31) between strings s_i^m and s_j^m, and the sum is over a set of meanings M of magnitude $|M|$.

Measure of Structure. For a particular language, a measure of structure is computed as follows. The distances between all pairs of strings in the language are calculated using normalized Levenshtein distance. In addition, the distances between all pairs of meanings also are calculated using a simple hamming distance (so that meanings differing in 1 feature have a distance of 1, meanings differing in 2 features have a distance of 2, and so forth). The Pearson's product-moment correlation between these 2 sets of distances then is calculated, giving an indication of the extent to which similar meanings are expressed using similar strings. To compare across different languages and to measure significance, it is necessary to compute a Monte Carlo sample of this measure under permutations of the strings over meanings. The graphs shown in the paper give the z score for the veridical correlation based on 1,000 randomizations. The dotted line on the graph therefore shows the 95% confidence interval that the observed mapping could be obtained by random assignment of signals to meanings. This measure is undefined when there is no variation in the Monte Carlo sample, for example when the language has only the same string for all meanings or for all but 1 of the meanings. In these cases, all possible reorderings are equally structured.

ACKNOWLEDGMENTS. H.C. was supported by an Economic and Social Research Council 1 + 3 Studentship. K.S. was supported by a British Academy postdoctoral fellowship.

1. Maynard Smith J, Szathmary E (1997) *The Major Transitions in Evolution* (Oxford Univ Press, Oxford, UK).
2. Deacon TW (1997) *The Symbolic Species: The Co-evolution of Language and the Brain* (WW Norton, New York).
3. Christiansen MH, Chater N (2008) Language as shaped by the brain. *Behav Brain Sci*, in press.
4. Kirby S (1999) *Function, Selection and Innateness: The Emergence of Language Universals* (Oxford Univ Press, Oxford UK).
5. Hurford JR (2000) Social transmission favours linguistic generalization. *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, eds Knight C, Studdert-Kennedy M, Hurford JR (Cambridge Univ Press, Cambridge, UK), pp 324–352.
6. Kirby S, Smith K, Brighton H (2004) From UG to universals: Linguistic adaptation through iterated learning. *Stud Lang* 28:587–607.
7. Kirby S, Dowman M, Griffiths T (2007) Innateness and culture in the evolution of language. *Proc Natl Acad Sci USA* 104:5241–5245.
8. Kirby S (2000) Syntax without natural selection, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, eds Knight C, Studdert-Kennedy M, Hurford JR (Cambridge Univ Press, Cambridge, UK), pp 303–323.
9. Brighton H (2002) Compositional syntax from cultural transmission. *Artif Life* 8:25–54.
10. Zuidema W (2003) How the poverty of the stimulus solves the poverty of the stimulus, *Advances in Neural Information Processing Systems 15 (Proceedings of NIPS'02)*, eds Becker S, Thrun S, Obermayer K (MIT Press, Cambridge, MA), pp 51–58.
11. Smith K, Brighton H, Kirby S (2003) Complex systems in language evolution: The cultural emergence of compositional structure. *Adv Complex Syst* 6:537–558.
12. Brighton H, Smith K, Kirby S (2005) Language as an evolutionary system. *Physics of Life Reviews* 2:177–226.
13. Griffiths TL, Kalish ML (2007) Language evolution by iterated learning with Bayesian agents. *Cogn Sci* 31:441–480.
14. Keller R (1994) *On Language Change: The Invisible Hand in Language* (Routledge, London).
15. Batali J (2002) The negotiation and acquisition of recursive grammars as a result of competition among exemplars, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, ed Briscoe EJ (Cambridge Univ Press, Cambridge, UK), pp 111–172.
16. Bickerton D (2007) Language evolution: a brief guide for linguists. *Lingua* 117:510–526.
17. Bartlett FC (1932) *On Remembering* (Macmillan, Oxford).
18. Curio E, Ulrich E, Vieth W (1978) Cultural transmission of enemy recognition: One function of avian mobbing. *Science* 202:899–901.
19. Laland KN, Plotkin HC (1990) Social learning and social transmission of foraging information in Norway rats (*Rattus norvegicus*). *Anim Learn Behav* 18:246–251.
20. Galef BG, Jr., Allen C (1995) A new model system for studying animal traditions. *Anim Behav* 50:705–717.
21. Laland KN, Williams K (1997) Shoaling generates social learning of foraging information in guppies. *Anim Behav* 53:1161–1169.
22. Mesoudi A, Whiten A (2004) The hierarchical transformation of event knowledge in human cultural transmission. *J Cognit Cult* 4:1–24.
23. Mesoudi A, Whiten A, Dunbar RIM (2006) A bias for social information in human cultural transmission. *Br J Psychol* 97:405–423.
24. Horner V, Whiten A, Flynn E, de Waal FBM (2006) Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proc Natl Acad Sci USA* 103:13878–13883.
25. Gomez RL, Gerken L (2000) Infant artificial language learning and language acquisition. *Trends Cogn Sci* 4:178–186.
26. Hudson-Kam C, Newport E (2005) Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Lang Learn Dev* 1:151–195.
27. Galantucci B (2005) An experimental study of the emergence of human communication systems. *Cogn Sci* 29:737–767.
28. Garrod S, Doherty G (1994) Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition* 53:181–215.
29. Garrod S, Fay N, Lee J, Oberlander J, MacLeod T (2007) Foundations of representation: Where might graphical symbol systems come from? *Cogn Sci* 31:961–987.
30. Selten R, Warglien M (2007) The emergence of simple languages in an experimental coordination game. *Proc Natl Acad Sci USA* 104:7361–7366.
31. Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions and reversals. *Sov Phys Dokl* 6:707–710.
32. Landau B, Smith LB, Jones SS (1988) The importance of shape in early lexical learning. *Cogn Dev* 3:299–321.
33. Slobin DI (1977) in *Language Learning and Thought*, Macnamara J ed (Academic, London), pp 185–221.
34. Smith K (2004) The evolution of vocabulary. *J Theor Biol* 228:127–142.
35. Von Frisch K (1974) Decoding the language of the bee. *Science* 185:663–668.
36. Zuberbühler K (2002) A syntactic rule in forest monkey communication. *Anim Behav* 63:293–299.

Language, Learning and Cultural Evolution: how linguistic transmission leads to cumulative adaptation

Simon Kirby*[1], Kenny Smith† & Hannah Cornish*
*Language Evolution and Computation Research Unit, University of Edinburgh
†Cognition and Communication Research Centre,Northumbria University

1. Introduction

An explanatory approach to language must, among other things, answer the question why language is structured in the particular way it is and not some other way. In other words, we seek to account for the particular universal properties of linguistic structure. Attempts to tackle this challenge take many forms (Hawkins, 1988), but in this chapter we look at a particular type of explanation, which we can term the *adaptive systems approach*.

This approach to an explanatory account for language focusses on its dynamical aspects, noting that the universal properties of language are actually the result of multiple complex dynamical systems operating on different time-scales each influencing the others. Specifically:

- **Learning/use.** The language produced by an individual is shaped in part by the cognitive mechanisms for learning and processing language. In other words, an individual's language adapts on an ontogenetic time-scale through acquisition and use.
- **Cultural[2] evolution.** The actual language spoken by any individual is also, obviously, a result of the language spoken by other individuals in the community and goes on to affect the language of future generations of speakers. Language universals arise from the interaction of individuals with particular cognitive and usage-based constraints in populations who share language. To put it another way, language is transmitted through a repeated cycle of learning and use leading to a process of change and evolution on a cultural time-scale (e.g., Brighton et al, 2005).
- **Biological evolution.** Finally, the cognitive machinery that drives the cultural evolution of language is itself the result of biological evolution. This leads to the possibility that the universals that emerge through cultural evolution may alter the fitness landscape of the individuals that learn and use these languages, ultimately leading to the biological evolution of the mechanisms for learning and processing language (e.g., Briscoe, 2000).

When we talk about these systems as being *adaptive* we mean that they result in the "appearance of design". That is, there is a fit between the structure that is the result of the dynamical system and some function of that structure. Adaptation is most familiar in the context of biological evolution, where natural selection is often seen as an optimising process generating phenotypes that are fit

[1] Author to whom correspondence should be addressed.

[2] The use of the term *culture* is potentially problematic. Throughout this chapter we use it in a technical sense to mean any information that is transmitted through a population by means of production of behaviour by an individual and acquisition of similar behaviour by another individual through observation. Language is transmitted culturally in this sense, but this does not mean it is necessarily shaped by other aspects of the "culture" of the individuals that posses it.

for survival and reproduction, but our point is that this is only one example of possible adaptive mechanisms.

At the core of this multiple adaptive systems approach to language is the idea that a) much of language structure is adaptive and b) whilst *appearing* to be designed there is no actual designer involved. This chapter will look mainly at the latter claim with respect to cultural evolution in particular by reference to mathematical, computational and experimental models of the transmission of language. Briefly, we aim to show that the process of transmission of language through repeated acquisition and use leads to cumulative adaptations without the need for biological evolution or any intention to adapt language on the part of those that use it.

2. The orthodox evolutionary view

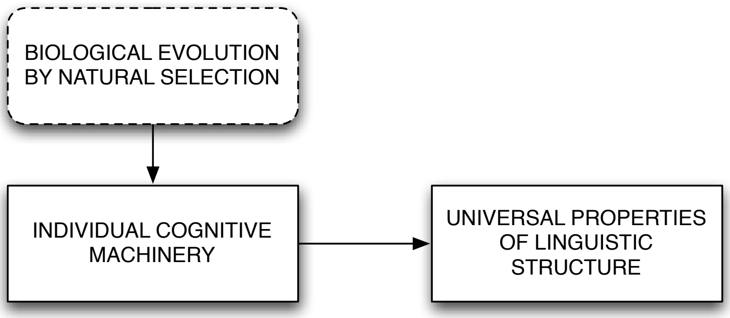


Figure 1*: The orthodox evolutionary view. The universal properties of linguistic structure are determined by the nature of our individual cognitive machinery which is the result of biological evolution under natural selection for communication.*

Faced with explaining the universal properties of linguistic structure, one influential approach has been a direct appeal to biology. In this view, language structure arises from our species-specific biological endowment – we have the types of languages that we do because we have an innately-given language faculty with a particular structure that constrains the possible types of language (e.g., Hoekstra & Kooij, 1988). In particular, Chomsky (1975) suggests that it is a set of innate constraints on language acquisition that determines the nature of human language.

For many (e.g. Hurford, 1990) this is an unsatisfying explanation as it stands, since it appears simply to push the need for answers back but not dispel them. It transforms one puzzle "why do we have the particular language universals we do?" into another "why do we have the particular language faculty we do?". In a landmark paper Pinker & Bloom (1990) set out a strategy for answering this second question in order to support a broadly nativist approach to explanation. This strategy has become what might be called the orthodox evolutionary approach to language (see figure 1).

Pinker and Bloom (1990) argue that language structure has all the hallmarks of an adaptation. To them, many of the fundamental features of language appear to be tailored to communicating complex propositions through a serial signalling medium. If Chomsky is right in arguing that these features of language are the way they are because they arise from an innately given faculty for language, then this makes language appear like many other features of our biology. The language faculty, like the heart or the liver, is an organ that appears adapted to a particular survival-relevant function – in this case communication.

If this is correct then the structure of the language faculty, like the structure of other organs, is best explained by appealing to biological evolution by natural selection. As they put it:

> "Grammar is a complex mechanism tailored to the transmission of propositional structures through a serial interface... Evolutionary theory offers clear criteria for when a trait should be attributed to natural selection: complex design for some function, and the absence of alternative processes capable of explaining such complexity. Human language meets this criterion." (Pinker & Bloom, 1990:707)

This biological/evolutionary approach to linguistic explanation is appealing since it neatly grounds out the explanation of linguistic structure in the well-established mechanism of natural selection.

Despite its appeal, there are reasons to be cautious with this orthodox evolutionary approach as it stands. One problem with the view portrayed in figure 1 is the link between "individual cognitive machinery" and "universal properties of linguistic structure". The Chomskyan approach to explaining language universals rests on a tacit assumption that constraints/biases on language acquisition will directly lead to equivalent constraints/biases on the distribution of possible human languages. But is this assumption justified?

A lesson can be learned from a different way of explaining language universals known as the *functional/typological* approach. Here, universals are explained by appealing not to innate characteristics of our language acquisition machinery, but rather to properties of the uses language is put to. We will not be looking at this literature in any detail here, but one of the criticisms levelled at it is that it fails to solve what has been termed *the problem of linkage*: how exactly does a feature of language use end up being reflected in the cross-linguistic distribution of language types (Kirby, 1999)? The point is not that this problem is insoluble, but rather it is an absolutely crucial part of any explanation. What is the mechanism that links the proposed *explanans* to the *explanandum* in question?

This linkage problem exists just as forcefully for the Chomskyan approach (see Kirby et al, 2004 for discussion):

> **Problem of Linkage.** Given a set of observed constraints on cross-linguistic variation, and a corresponding pattern of functional preference *or language acquisition biases*, an explanation of this fit will solve the problem: how does the latter give rise to the former? [Italic text added to the original definition from Kirby, 1999]

What is needed is a way of bridging the gap between an individual-level phenomenon (the structure of a language-learner's cognitive machinery) and a population-level phenomenon (the distribution of possible languages). As Kirby et al (2004) argue, the solution to this problem is to explicitly model the way in which individual behaviour leads to population effects over time. As noted in the introduction, language emerges out of a repeated cycle of language learning and language use, and it is by studying this socio-cultural process directly that we will see how properties of the individual leave their mark on the universal structure of language.

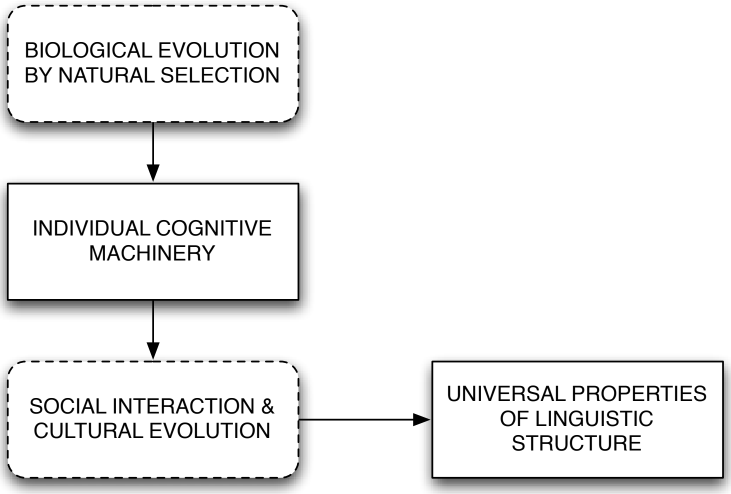


Figure 2: *The solution to the Problem of Linkage. The universal properties of language arise from the cultural evolution of language through generations of socially interacting individuals with particular cognitive machinery. Of central importance is the precise contribution of the cultural evolutionary process in determining language structure.*

Of course, it may well be that when we examine this linking mechanism we will find that language universals do indeed straightforwardly reflect language learning biases, for example. If this is the case, then the orthodox evolutionary explanation is a reasonable one. However if the extra box in figure 2 does some work for us, then this explanation cannot hold – at least in its present form. Indeed, we may find that the explanatory burden may be lifted to some extent from our innate machinery, and hence from biological adaptation through natural selection.

3. Modelling cultural evolution

So far we have identified the importance of understanding cultural evolution as it applies to language because it represents the solution to the problem of linkage in the orthodox explanation for linguistic structure. The difficulty is that we have a surprisingly poor understanding of exactly how cultural evolution actually works in general (although there is a growing literature, e.g. Boyd & Richerson, 1985, Mesoudi et al, 2006). Compared to our detailed empirical and theoretical understanding of language acquisition, for example, or the process of biological evolution by natural selection, we do not have a strong empirical base for cultural evolution or an accepted set of principles for how individual biases lead to population-level phenomena.

There is, of course, an extensive literature on historical linguistics which appears relevant. It is important to note that our target is subtly different. When linguists study language *change*, they consider how a language at one point in time turns into a different language at a later point. However, we would expect both of these languages to fall within the boundaries described by our theory of language universals. Normally, historical linguists are interested in how languages move around the space of possible languages rather than in the origins of that space in the first place.[3] We will return to this distinction later in a more formal context.

In order to better understand how cultural evolution works in general, and how it operates in shaping language in particular, we have set out to model it in three different ways over the past decade or so (given here in the order they have been explored):

- **Computational models.** Our first approach was to build simulations of populations of individuals with particular language learning machinery and see what types of languages emerge. The goal here was to examine the extent to which the resulting language structure was determined by features of the cultural transmission process rather than being directly encoded in the learning mechanisms (e.g. Kirby, 1994, 1999; Kirby & Hurford, 2002; Smith, 2002; Smith et al 2003;)
- **Mathematical models.** Based on our experience of the computational models, we developed an idealised mathematical framework which enabled us to state precisely how much our innate endowment determines the structure of language (e.g., Kirby, Dowman & Griffiths, 2007).
- **Experimental models.** Finally, to act as a check on the plausibility of the formal models and to see how closely human subjects behave like their computational idealisations, we developed a novel experimental paradigm for cultural evolution (e.g., Cornish, 2006).

All three of these are based on a framework for understanding cultural evolution we have called the *iterated learning model* (see figure 3). Iterated learning is the fundamental process underlying many forms of cultural evolution, including language. It is the process of the transmission of behaviour where that behaviour is acquired by an individual observing similar behaviour in another who acquired it in the same way. The model, based on Andersen's (1973) and Hurford's (1990) treats the transmission of language as a repeated transformation between some linguistic representation internal to an individual (or "agent" to use the modeller's parlance) and utterances that are external to that individual and can be observed by another. It is through being repeatedly learned and used by agents in the model that language evolves culturally.

Because our aim here is not a theory of language change, we do not typically start the models off with something that falls within the space of possible human languages. Instead we are interested in how (and whether) such human-like languages emerge in the models when one is not present in the initial conditions (see Brighton, 2003, for a detailed discussion of the methodological issues this raises). By varying features of the way in which language is transmitted from agent to agent in the models, we can begin to build-up a picture of how cultural evolution might work.

For the remainder of this chapter, we will briefly review the main results so far from the three strands of modelling research listed above and discuss what they tell us about how we should approach linguistic explanation. Of particular interest will be the question: how much of language structure that appears to be designed for communication need not be explained in terms of the intentions of communicating agents at all?

[3] There are some exceptions to this. For example, there have been attempts to apply grammaticalisation theory to the origins of language by reconstructing a pre-existing state where language universals would have been different (Heine & Kuteva, 2002).

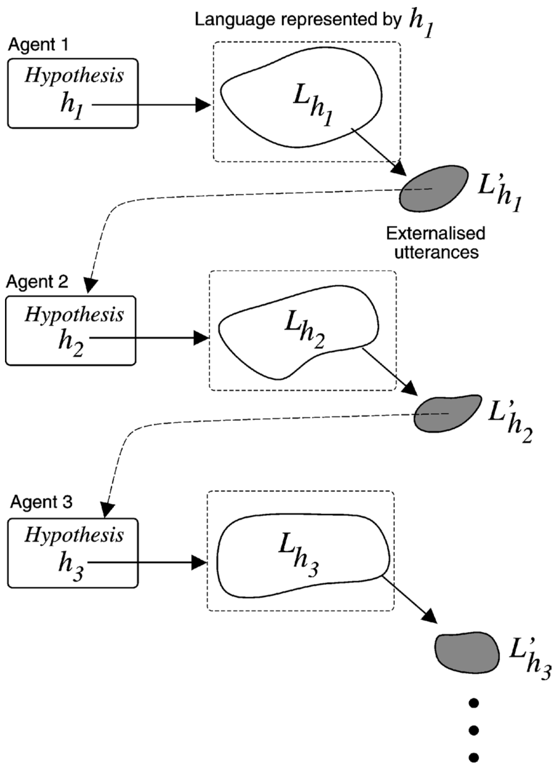


Figure 3: *The Iterated Learning Model. The first agent has knowledge of language represented by a hypothesis h_1 . This hypothesis itself represents a language L_{h1} . Some subset of this mapping, L'_{h1}, is externalized as linguistic performance for the next agent to learn from. The process of learning results in a hypothesis h_2. The process is then repeated, generation after generation.(Taken from Brighton, et al, 2005:185)*

4. Computational models: language transmission is adaptive

Since the early nineties, there have been a variety of attempts by a number of researchers (e.g. Kirby, 1994; Batali, 1998; Kirby, 2002a; Teal & Taylor, 1999; Tonkes, 2001; K. Smith, 2002; Steels et al, 2002; Brighton, 2003; Zuidema, 2003; de Boer, 2005; A. Smith 2005; Vogt, 2005; Oudeyer, 2006) to build simulation models of the cultural evolution of language (see, e.g. Kirby, 2002b; Steels, 2003, for review and Brighton et al, 2005, for a detailed account of one particular strand of research). Most of these models adopt a framework similar to the one outlined in figure 3: a population of agents produce language-like behaviour in response to observing similar behaviour in other members of the population. They differ (often radically) in their assumptions about the nature of the population, their model of learning, and exactly what form the agents' language takes.

For example, Batali (2002) has a model in which there is a relatively large but static population of agents throughout the simulation, whereas Kirby (2000) implements a model with purely vertical cultural transmission in a chain of "adults" and "children". In the former, no-one is born and no-one dies, but in the latter there is strict generational turnover with the children replacing the adults each generation and a new set of children being introduced. In Batali (1998) agents are recurrent neural networks being trained using standard connectionist algorithms, whereas Brighton's (2002) agents induce finite state machines using Minimum Description Length learning. In Kirby (2002a), agents communicate meanings represented as hierarchical symbolic propositions, whereas Vogt's (2005) agents play communication games grounded in visual stimuli.

Despite the large range of different assumptions, methods and motivations across these models one broad conclusion seems warranted (cf. Christiansen, 1994; Deacon, 1997):

> **The Principle of Linguistic Adaptation:** if a learner is given imperfect information about the language they are attempting to learn (e.g., if they are subject to noise, processing constraints, or they simply do not hear all the data), then cultural transmission becomes an adaptive system. As a result, languages will emerge that appear to be optimised to the problem of being transmitted from individual to individual.

We can think of the transmission of the knowledge of language from one agent to another as passing through a narrow "bottleneck". A large (or potentially infinite) language must be reconstructed by a learner despite the imperfect information imposed by the bottleneck.[4] The very act of repeatedly squeezing language through this bottleneck causes language to change in such a way that its chance of being transmitted through the bottleneck with high fidelity is maximised.

The literature provides numerous examples of the principle of linguistic adaptation at work in simulations and shows how it can be used to cast light on specific linguistic problems. Here we provide three illustrations from our own work in this area. These summaries are necessarily brief, but nevertheless we hope they will give a flavour of the work in this area.

Hierarchical universals and competing motivations

Language universals of the sort discussed in the typological literature are often implicational in nature (e.g., Croft, 1990). That is, languages are predicted to have property Q if they also have property P, but not necessarily vice versa. In other words: P→Q. In some cases, researchers have uncovered whole chains of implications of the form P→Q & Q→R & R→S etc. These are often rewritten in the form of hierarchies of types: S>R>Q>P. Languages which exhibit a feature at some point in the hierarchy will also exhibit all the features higher in the hierarchy.

An influential typological study of relative clause formation provides a prototypical study of hierarchical universals. Keenan & Comrie (1977) present evidence for the following hierarchy of accessibility to relative clause formation:

> **The Relative Clause Accessibility Hierarchy:**
> Subject>Direct Object>Indirect Object>Oblique>Genitive>Object of Comparison
> If a language can relativize any position on this hierarchy, they can relativise all higher positions in the hierarchy.

How might we explain a universal pattern such as this one? Kirby (1997) examines an explanation due to Hawkins (1994) that appeals to asymmetries in the difficulty in processing different relative clauses. Simplifying somewhat, the idea is that the greater the structural distance between a head noun and the trace (or resumptive pronoun) in a relative clause, the greater the load there is on the working memory of the parser. However, this fails to explain why there should be a link

[4] Note that this way of expressing things suggests an analysis of this sort of cultural evolution along the lines of Sperber (1996). However, it is also possible that the evolution of language in these models could be studied in terms of populations of competing replicators (e.g. Croft 2000, Kirby 1999).

between processing load during parsing and the observed universal – what is the mechanism that links the two?

In order to solve the problem of linkage here, Kirby (1997) sets out to model, in a simple simulation, how this parsing preference actually results in a hierarchical universal.[5] The simulation consists of a population of agents. Each agent has a grammar which either allows or disallows relative clauses for each point on the hierarchy. The population of agents is updated through a process of generational turnover whereby a new population of agents is created; each agent in the previous generation produces example relative clauses according to their grammar; and the new agents acquire their grammars on the basis of the examples produced by the agents from the previous generation. The learning mechanism is set up in such a way that the probability that a learner acquires a particular relative clause type is dependent both on the number of examples the learner hears and the parsing difficulty associated with each clause type. This implements in a straightforward way a parsing-based bottleneck on the cultural transmission of language.

The results of this model immediately demonstrate a problem with the explanation for Keenan and Comrie's (1977) hierarchy as it stands. No matter what the initial distribution of languages is in the population, the only stable end state is one where languages don't allow any relative clauses at all. It is easy to see why this happens: language is simply adapting to the complexity of processing relative clauses. The most adapted languages are those that avoid the problem of relative clause processing by rendering them ungrammatical.

Kirby (1997) shows that this is a general problem with any explanation for hierarchical universals that appeals to an asymmetry in processing difficulty. The solution, verified by the simulation model, is to seek a *competing functional motivation* favouring the structure in question. Interestingly, this competing motivation need not be asymmetrical with respect to the different types on the hierarchy. So, all that is needed to derive the observed distribution of language types in the simulation is a general speaker-driven least-effort principle that *favours* relative clauses of all types equally, for example because they avoid the need for circumlocution. With this pressure acting on speakers (and some assumptions about how the relative strengths of pressures may vary over time) the end result of the simulation is a distribution of languages that obeys the Keenan and Comrie hierarchy. All the language types at the start of the simulation that do not correspond to those found in the world today disappear.

What these results demonstrate is that languages can adapt to competing needs of speakers and hearers as they influence the bottleneck on linguistic transmission. The hierarchy is not built-in directly as a set of constraints on possible languages – nor do the agents in any way try and optimise the language they have. The universal emerges as a population-level effect from processing pressures acting on individuals influencing the transmission of language through iterated learning.

Compositionality and morphological regularity

Whereas the early iterated learning models looked at specific language universals of the sort uncovered by typological surveys, with steadily increasing computing power and interest in the evolution of language there has been a desire among many researchers to simulate the emergence of

---

[5] Although this explanation appeals to parsing preferences that are arguably innately given, it appears to be of a rather different type than the Chomskyan style of explanation from innate language acquisition constraints outlined earlier. Indeed the problem of linkage referred to here is essentially the one that pervades much functional explanation of typological generalisations. Nevertheless, as we shall see, exactly the same linkage problem exists for explanations appealing to acquisition constraints/biases, and the solution to the problem is the same as well.

language out of a pre-existing a-lingual state. For example, could some of the fundamental features of syntax be shown to evolve from a largely non-syntactic protolanguage[6] solely through cultural processes?

A particularly fundamental structural feature of language that sets it apart from almost all other communication systems in nature is *compositionality.* It is regular compositionality in the mapping between signals and meanings that, when recursively applied, gives language its completely open-ended expressivity. Drawing on a variety of evidence, Wray (1998) proposes an earlier stage in the evolution of language where signals and meanings are not related compositionally, but rather whole signals correspond to whole meanings. This *holistic protolanguage* is in many ways closer to the communication systems of non-human primates, which are based on a fixed repertoire of expressions lacking generalisable internal structure.

The puzzle is what drives the transition from a holistic stage in language to a more syntactic system of communication. Why and how does compositionality emerge? Can the principle of linguistic adaptation help?

If we think about the difference between holistic and compositional mappings from the point of view of the transmission of language, it becomes obvious that the principle of linguistic adaptation does indeed predict that compositionality will emerge in most cases. Assuming that there is a larger range of meanings that an individual language learner *could* be exposed to in their lifetime than the range of meanings that they actually *are* exposed to, then there is a bottleneck on linguistic transmission because a learner will never see the entire language. This means that a holistic expression for some meaning will only ever be learned if that exact expression is observed by a learner. On the other hand, in a compositional language, a sub-expression (e.g. a word or morpheme) corresponding to a sub-part of a meaning has a much greater opportunity to be learned since evidence for it can be seen by a learner whenever any meaning in which it is involved is expressed. Hence, generalisable linguistic structure is better able to fit through the bottleneck on linguistic transmission. Jim Hurford puts it succinctly in the title to his article: "social transmission favours linguistic generalisation" (Hurford 2000).

Kirby (2001) demonstrates the process at work in a computational simulation. Agents in this model acquire languages from observations of strings of characters being paired with a finite set of very simple structured meanings. Meanings are essentially pairs of features, each of which can take a range of values. The initial expressions in the simulation are random strings of characters paired with whole meanings. In other words, the initial language is holistic because there is no regularity in the mapping between meanings and signals.

Agents are prompted to produce signals for meanings at random and will do so using their internalised language if possible, otherwise they will "invent" a random novel string of characters if necessary. Learners store signal-meaning pairs that they hear in a list, but will also search for any generalisations they can make over the set of pairs that they store. Of course, given a purely holistic language there are no generalisations that can be made, so the language remains holistic.

What happens in such a model? It turns out that it depends critically on how much data learners see in their lifetime. As predicted, this learning bottleneck drives the cultural evolution of language as it is transmitted from generation to generation in the iterated learning model. When learners see large amounts of data, then the language typically is acquired perfectly each generation and therefore does not change. In this case, a completely holistic protolanguage is stable. However, if the number of meaning-signal pairs each learner is exposed to is reduced then the language becomes

[6] We use *protolanguage* here in its evolutionary sense (e.g. Bickerton, 1990; Wray, 1998) to mean an evolutionarily prior form of language without all the hallmarks of modern human language.

rather unstable. This is simply because agents will be called upon to produce signals for meanings they have never encountered in their input. Because the language is holistic, their only option is to produce a novel random string. The particular meanings that are subject to this random innovation differ each generation in the simulation (because meanings are picked at random for agents to produce). The upshot of this is that the language can change from generation to generation.

If this were all that happened, it would not be a very interesting model. However, something rather striking occurs when there is a learning bottleneck such as this one: the initially unstable language transforms over time into one that is stable despite, or rather *because of*, the limited input to learners. This new stable language is compositional. Each feature ends up being expressed by some sub-part of the signal. So, for example, a complete meaning might be encoded by using a "morpheme" corresponding to the value of the first feature attached to a "morpheme" corresponding to the value of the second feature.

This compositional coding system emerges piecemeal (but surprisingly rapidly) in this simulation as speakers' purely random and holistic innovations are incorrectly over-generalised by learners. The crucial point that arises from the iterated learning model is that these mistaken over-generalisations are then *correctly* picked-up by learners in the next generation. Because generalisations are better able to get through the learning bottleneck, this process snowballs and the inevitable end-result is the emergence of rampant compositionality.

It is important to realise that this result is not simply an artefact of particular features of this one simulation. As noted in the introduction to this section, the same basic behaviour can be seen in simulation models with radically different assumptions and architectures. Furthermore, this type of model can not only provide an explanation for the origins of compositional regularity, but also explain the cases where it does not occur. Whilst most simulation models make the simplifying assumption that all meanings were equally frequent, Kirby (2001) implemented a non-uniform frequency distribution in his model so that some combinations of feature-values were more likely to be expressed by speakers than others.

	a_0	a_1	a_2	a_3	a_4
b_0	**g**	**s**	kf	jf	uhlf
b_1	**y**	jgi	ki	ji	uhli
b_2	yq	jgq	kq	jq	uhlq
b_3	ybq	jgbq	kbq	jbq	uhlbq
b_4	yuqeg	jguqeg	kuqeg	juqeg	uhluqeg

Figure 4: *Simulation result showing a partially regular paradigm. Meanings involve two components, "a" and "b". Frequency of these combinations increases to the upper left of the table. The signals are combinations of letters and exhibit regular compositional structure except for the most frequent meanings. (Taken from Kirby 2001)*

In this case, only infrequent meanings end up being expressed compositionally. Highly frequent meanings tended to remain with irregular holistic forms (see figure 4). This makes sense from the point of view of the principle of linguistic adaptation. If a meaning crops up with high frequency, then information about how that meaning is expressed is reliably provided to the learners. There is no pressure in this case for it to be regularised and become compositional. This result is suggestive in the light of the well-known relationship between frequency and regularity in the morphology of real languages. For example, the top ten verbs in English by frequency all have irregular past-tenses (Francis & Kucera, 1982).

The adaptation of meanings through iterated learning

Both of the previous examples show how the *form* of language may adapt through a process of cultural evolution. In the second example, the structure of strings in the language end up largely mirroring the pre-existing structure of the meanings in the model. In these simulations, the meaning structure is defined and fixed by the experimenter, leading some researchers to wonder if more flexible meanings can be modelled in simulation (e.g, Steels et al, 2002; A. Smith, 2005; Vogt, 2005).

Indeed, in a simple idealised computational model, Kirby (2007) suggests that semantics as well as syntax might adapt through a process of cultural evolution under pressure from a bottleneck on transmission. To model this, a distinction is made between the meanings that the agents associate with signals on the one hand, and their actual communicative goals on the other. So, for example, I as a speaker may wish to draw the attention of a hearer to a particular person in a room. I might choose to do this in a number of ways: from describing them in every detail, through simply noting their distinctive features, to referring to them by name. In the computational model, each of these correspond to different "meanings" associated with the object of reference. In some sense, it is up to the speaker to choose which meaning they wish to convey, which in turn will affect the actual signal produced. Note that, at one extreme this corresponds to a holistic system of communication – in their simplest form, proper names are holistic. At the other extreme, we might imagine a deeply compositional (but highly inefficient!) form of communication where every discernible aspect of the object of reference is explicitly expressed.

For brevity, we will omit the details of the model here, but the key is that although agents are able to conceptualise every communicative goal in a large number of different ways, corresponding to different meanings, they are only able to express those meanings if they have previously encountered similar expressions in their training data. More precisely, they are able to express a target meaning if they have previously heard a set of meanings within which all aspects of the target meaning appear at least once. To put it another way, just as in the previous model, compositionality allows the learners to recombine sub-parts of other expressions to form novel ones as long as there is sufficient evidence in the input.

Agents are randomly given a particular communicative goal and a "context" of a number of other randomly chosen irrelevant goals. They then try and find a meaning corresponding to their communicative goal for which a suitable expression can be generated. If more than one meaning is possible, then agents pick one which best discriminates the target of communication from the context. If no meanings are possible, then as in the previous model, agents invent a new expression.

The result of these modifications to the previous iterated learning models is that the kinds of meanings that agents use evolves culturally, rather than simply the signals that they associate with meanings. So far the analysis of the model is far from complete, but what is clear is that the language once again shows evidence of adaptation. Where there is pressure from the learning bottleneck, meanings are preferred which allow for the most generalisable forms of compositional language.

There are a number of flaws in this model, unfortunately. For example, although the use of meanings can change over time the set of all possible meanings must still be provided somehow by the experimenter. Work by robotics and artificial life researchers looking at the origins of communication may eventually provide the best way out of this issue by grounding communication in the real world (e.g. Steels, 2003) or some model of ecological relevance (e.g. Cangelosi et al, 2002). However, the important point here is to show that principle of linguistic adaptation may potentially have a very wide remit in helping explain many aspects of linguistic structure.[7]

5. A mathematical model: from weak innateness to strong universals

The computational models reviewed briefly in the previous section lend credence to the notion that language is an adaptive system in its own right. Features of the bottleneck on linguistic transmission end up influencing the structure of language as it adapts through a process of cultural evolution to the challenge of being repeatedly learned by generations of agents.

It is worth reviewing at this point the relevance of linguistic adaptation to the nativist argument outlined in section 2. We highlighted the importance of tackling the problem of linkage when considering nativist explanations. We proposed that cultural evolution is the mechanism that links properties of an individual's language learning machinery with universal features of linguistic structure. If the result of cultural evolution is a straightforward expression of innate biases in cross-linguistic distribution (i.e. if Universal Grammar gets expressed directly as language universals), then there is no particular problem with the orthodox evolutionary view. However, another possibility is that the contribution of cultural evolution is more significant – that it distorts or transforms the innate biases in such a way that their explanatory significance is reduced.

It certainly seems likely that the latter is true given the results of the computational models. For example, it is clear that features such as the amount of training data and the frequency of meanings have significant (and even determining) influence on fundamental features of the structure of the languages that emerge. That said, there are problems with the simulation models as they stand.

Most crucially, it is very difficult to say for any given computational model exactly what the contribution of innate biases actually is, or even what those biases are in the model. The fact that similar results are achieved with hugely different architectures suggest that whatever prior biases the models have (and they surely have some since bias-free learning is impossible) their details might not have a strong bearing on the outcome. It would nevertheless be nice if we could know exactly what the relationship between innateness and universals is in general and it is hard to see how this kind of simulation model is going to be able to do that.

To tackle this question, we can use a general model of learning which makes prior bias explicit and embed this in a mathematical idealisation of the iterated learning model (Kirby, Dowman & Griffiths, 2007, building on Griffiths & Kalish, 2005). This model treats learning as a process of selecting the best hypothesis (i.e. grammar) given a set of data (i.e. utterances) and a prior bias towards some hypotheses over others (i.e. a model of innateness). Bayes' law provides us with a neat mathematical characterisation of how these interact. We can use it to calculate the probability of a hypothesis given some data (which is what a learner would ideally like to know) from the probability of the data given that hypothesis (which can be estimated if we know how utterances are produced) and the prior probability of that hypothesis independent of any data seen (which is the innate contribution of the learner's machinery):

$$p(h \mid d) \propto p(d \mid h)\, p(h)$$

where h is the hypothesis under question and d is the set of data heard by the learner.

---

[7] The brief overview given here is far from exhaustive in this regard too. For example, work by Oudeyer (2006) demonstrates how similar ideas can explain the origins of the phonemic code.

If we assume that learners pick the best hypothesis they can – the one that maximises $p(h \mid d)$ – then we can in principle construct a complete view of the dynamics of iterated learning for any model of hypothesis space, innate contribution and production model. We simply calculate for any pair of hypotheses (i.e. languages), h_i and h_j, the probability that a speaker with hypothesis h_i will produce data that a learner will infer has actually been produced by h_j.

This set of probabilities defines a transition matrix over languages (c.f. the Q-matrix of Nowak et al, 2002) showing how languages will change over time as they are repeatedly used and acquired. It turns out that there are straightforward mathematical techniques for transforming such a matrix into a probability distribution over languages corresponding to the predicted cross-linguistic distribution as an outcome of iterated learning.[8]

What this set of mathematical tools gives us is a way of plugging-in different assumptions about innateness and seeing exactly how they result in language universals. Kirby, Dowman & Griffiths (2007) use this to test whether the *strength* of innate biases is reflected in the resulting language universals. For example, are strong innate constraints required to explain big asymmetries in the distribution of language types? Equally, can the nature of innate biases be inferred straightforwardly from observed language universals?

Firstly, the mathematical results back-up the computational models in showing that languages adapt to the nature of the transmission bottleneck. Frequency of meanings and the number of examples learners are exposed to fundamentally shape the language universals that emerge. This in itself acts to obscure the influence of the prior bias (see, for example, figure 5 which confirms that frequent meaning are far more likely to be irregular despite a prior bias with only a slight and equal preference for regularity across the board).

A more striking result of this model, however, is that for a wide range of values, the actual strength of the prior bias makes *absolutely no difference* to the universal distribution that emerges. Although the nature of the prior bias is clearly important, the degree to which any innate preferences are reflected in the languages that emerge is dependent on such things as the number of examples seen rather than the strength of those innate preferences themselves.

[8] This distribution is called the *stationary distribution* and is the limiting distribution of the process of linguistic transmission (given some plausible assumptions about the nature of the transition matrix). The distribution is stationary in the sense that the *probabilities* of any particular language being found do not change, but the particular languages in a population at any point in time may. It can be thought of informally as the time average of languages after the dynamics of iterated learning have settled down. The stationary distribution gives us a way of thinking about the differences between the study of language change and the study of (cultural) language evolution mentioned in section 3. The former looks at how languages move within the stationary distribution. The latter looks at how the stationary distribution itself is formed.

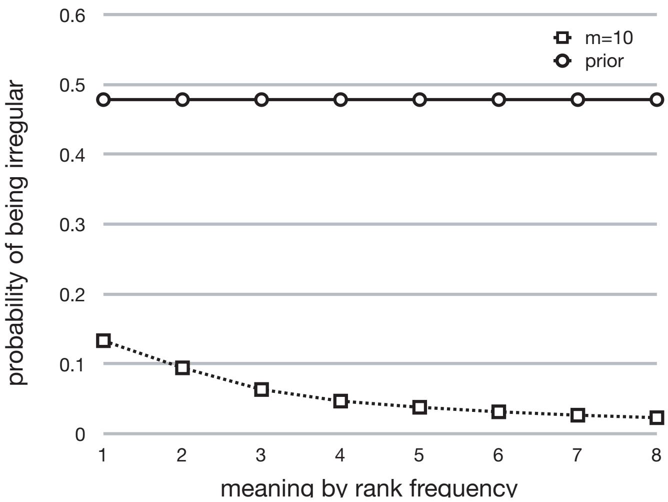


Figure 5: *The effect of cultural transmission on an innate bias. The graph shows the probability of a meanings in an abstract model of a language being irregular across meanings with decreasing frequency. The top line shows the expectation of irregularity encoded in the learners' prior learning bias. The lower line shows the actual probability of irregularity that emerges through cultural evolution. (See, Kirby, Dowman & Griffiths, 2007, for more details.)*

6. An experimental model: cumulative adaptation without intention

As we have seen, the results from a wide range of formal models lend credence to the principle of linguistic adaptation. Language appears to spontaneously adapt to pressures affecting its own survival through cultural transmission. Of course, there has been some scepticism of the validity of modelling results applied to the evolution of language. For example, Bickerton (2007) complains:

> "Powerful and potentially interesting though this approach is, its failure to incorporate more realistic conditions (perhaps because these would be more difficult to simulate) sharply reduces any contribution it might make towards unravelling language evolution. So far, it is a classic case of looking for your car-keys where the street-lamps are." (Bickerton, 2007:522)

A pressing question is therefore, can this kind of cultural adaptation be observed in real human subjects rather than simulated agents?

Previous experimental work has looked at the emergence of novel systems of communication in groups of experimental subjects. For example, in experiments reminiscent of the board game *Pictionary* (Healey et al, 2002; Fay et al, 2004) show how subjects forced to communicate graphically (and disallowed from writing) can converge on a way of communicating meanings using graphical signals. A particularly fascinating result from this work is an observed transition from the use of icons to communicate in initial stages to a more symbolic mode of communication once shared conventions have been set up. In another experiment, Galantucci (2005) demonstrated that pairs of subjects could converge on a shared symbol system when trying to solve a cooperative computer game with only a graphical means of communication. Selten & Warglien (2007) demonstrate a similar process in a non-graphical paradigm.

These experiments focus on the role of *interactive feedback* in the creation of shared systems of communication. Participants attempt to construct a signalling system and modify it in response to feedback in the form of behaviour or further signalling by the other participant in the exchange. The systems that emerge are a product of a combination of deliberate design and interactive negociation between interlocutors.

But how good is this as a model of the emergence of linguistic structure? While it cannot be denied that humans are intentional agents who have the potential to design and construct communication systems – indeed these experiments demonstrate this admirably – it remains arguable that the universal structural properties of human languages are the result of such intentional design.

Keller (1994), for example, argues at length that much of human language is best seen as a result of an "invisible hand" process, echoing Smith's (1776) use of the term as a metaphor of the way individuals influence market economics. Keller's point is that language change, although being the result of actions of intentional agents, is not the goal of those agents' intentions. To put it baldly, the shift from OV to VO order in the history of English arose from the actions of speakers of the language and may ultimately have a functional motivation (Hawkins, 1994), but it was surely not the result of individuals deciding to modify the language in such a way to improve its parsing efficiency.

This kind of argument can be applied more broadly to cultural evolution. Many products of human behaviour are the result of intentional design, but some are non-intended consequences of many individuals' actions. It is possible that these non-intended consequences may nevertheless be adaptive – they may show the appearance of design without actually having a designer. The parallels here with biological evolution, where apparent design results not from an intentional Creator, but from the non-local consequence of local selection, has led some to propose studying cultural evolution in similar terms (e.g., Aunger, 2001).

Despite huge interest in these theories of cultural evolution, and their relevance to language evolution, there has been as far as we know no previous experimental validation that culturally transmitted behaviour can actually adapt without intentional design. Cornish (2006) and Kirby, Cornish & Smith (2008) set out to rectify this by setting up an analog of the computational models of iterated learning using human subjects rather than simulated agents.

These experiments combine two experimental paradigms: diffusion chain studies and artificial language learning. The former have previously been used among other things (e.g., Mesoudi et al 2006; Bartlett, 1932; Kalish et al, 2007) to look at whether chimpanzees are able to culturally transmit information about how to open a puzzle box (Horner at al, 2006). A chain of experimental subjects is set up in which each one observes the performance of the previous subject in the experiment and then in turn produces behaviour that the next subject is able to observe. In this way, the task that a subject faces is in some sense outwith the experimenters control (excepting the initial participant in the experiment) because it is ultimately determined by the previous participants' behaviours. The (perhaps surprising) result is that cultural transmission in chimpanzee populations has high fidelity – if a diffusion chain of chimpanzees is initialized with box opening behaviour A then that behaviour will be faithfully transmitted across a number of generations, without a switch to the equally functional opening behaviour B.

Artificial language learning experiments, on the other hand, examine the performance of individuals at learning a particular hand-constructed artificial language (e.g., Gomez & Gerken, 2000) with a goal of determining human language learning biases. We can think of this paradigm as mirroring the behaviour of one individual agent in the iterated learning simulations, whereas the diffusion chain experiments are akin to the population model where behaviours are repeatedly transmitted from "adults" to "children".

The experiment

Subjects in our experiment are treated as if they were participating in a standard artificial language learning task. They are told they are going to be exposed to an "alien" language that they must try and learn. The experiment starts with a random, unstructured, holistic language which is presented to the first subject. After training on this language, the subject is tested and their output recorded. The innovation is the embedding of this task within a diffusion chain: the output of the first experimental subject forms the language that the second subject in the experiment will be trained on, and so on. In this way, we can track how the initially random language changes as it is repeatedly learned and produced by "generations" of participants in the experiment. Crucially, subjects are not aware of the cultural nature of the experiment. They are simply asked to give us back as best they can the language that we have presented to them. In other words, there is no sense that participants in the experiment are trying to improve the language in any way, for example to score well in some collaborative game.

The hypothesis being tested is that there will be cumulative cultural adaptation of the language without intentional design by participants. Accordingly, we expect two things to happen in experiments such as this one:

- the language should become easier to learn;
- and the language should become structured.

If this happens, then insofar as we can say that this was not the result of intentional design on the part of the participants in the experiment, the hypothesis will have been confirmed.

In order to make this kind of iterated learning experiment work, we need to have some way of eliciting language data from subjects. We cannot, for example, simply test subjects recall of strings in the input language through a forced-choice task. This is because we need to generate training data for the next participant. To get round this problem, we trained subjects on stimuli that were a combination of strings of written syllables and simple schematic pictures (corresponding to *signals* and *meanings* respectively). In the testing phase, subjects were asked to produce the correct string of syllables corresponding to each picture in turn, thus providing us with a new training set for the next generation.

Each picture/meaning in the experiment is a coloured shape moving in one of three ways (bouncing, spiralling or sliding). There are three possible shapes (square, circle or triangle) and three colours (red, blue or black), yielding 27 different meanings. The original language is constructed by randomly concatenating, without spaces, 2 to 4 CV syllables from a set of a possible 9. For example, in the original language, a red bouncing square might be labelled "kihemiwi". Although the initial language has these constraints, subjects are free to type any combination of characters they wish in their output at test.

At each generation (i.e., for each participant in the experiment), the input language is divided randomly into a SEEN and and UNSEEN set. Participants are trained a number of times on the SEEN set by being presented with each picture and string in turn on a computer screen. They are then *tested* on the entire SEEN + UNSEEN set of 27 pictures in order that we can gather a complete language. This new language is then divided (randomly again) into SEEN and UNSEEN sets for the next participant (see Cornish, 2006, for more details).

Results

The complete results and analysis of this experiment are given in Cornish (2006), but we give a brief summary here. In all cases our hypothesis of cumulative adaptation is confirmed.

Firstly, to see if learnability increases, we measured the difference in the strings produced by a subject at generation *n* with a subject at generation *n*-1. In the initial stages, the difference was extremely high. This is not at all surprising. After all, we are not only presenting subjects with a random set of strings, but we are asking them to respond with strings for pictures that they have never previously seen. For these unseen examples and the initial random language at least it is impossible to get these right (except by an overwhelming fluke!). Remarkably, however, as the experiment progressed, later generations found it increasingly easy to get strings correct, or near correct. In fact, in some cases after 7 or 8 generations had passed, subjects were getting *every* string correct even those for pictures they had never seen in the training data. In other words, the language evolves culturally to become more learnable.

How does it achieve this feat? Recall we predict that adaptation of the language should lead to structure evolving. This is indeed what happens, but the type of structure depends on how we divide up the data each generation into the SEEN and UNSEEN sets.

In our first experiment, the language was divided evenly into SEEN and UNSEEN, with 14 and 13 pictures in each respectively. The result was quite surprising given previous computational models, but makes a lot of sense in retrospect. The language adapts to be learnable primarily by reducing the number of distinct words. To put it another way, strings become ambiguous with respect to the pictures. The initial random language has 27 distinct words in it (one for each picture), but at the end of the experiment (which ran for 10 generations) the language only has 5 words. This alone does not capture everything that is going on, however, otherwise subjects would still not be able to get all UNSEEN pictures 100% correct as they do for the last three generations in the experiment. To do this, there must be some structure in the mapping from meanings to signals.

A statistical analysis of the language at each stage confirms that this structure exists. Basically, words end up being used for sets of pictures that tend to share features in common. The final language (which is stable for three generations) shows this most clearly:

- *miniku* is used for all bouncing circles
- *tupin* is used for all bouncing triangles
- *tupim* is used for all bouncing squares
- *poi* is used for anything that spirals
- *tuge* is used for anything that slides

What has happened is that word-picture pairs have been generalised in such a way that the language can pass through the learning bottleneck. Even if subjects do not see half of the pictures, they can nevertheless be reliably named.

This is not the kind of result familiar from the computational models reported earlier – structure in the signals themselves does not emerge. What seems to be missing here is any pressure on the language to be *expressive*. Languages with fewer words are clearly more learnable, so adaptation to learnability inevitably reduces the expressive power of the language by introducing what is essentially rampant polysemy and a reduction in the discriminative power of the language. Once the number of words is low enough no further adaptation is necessary since the language passes easily through the learning bottleneck.

For our second experiment, we made a minimal change to the procedure to try and reduce the amount of polysemy that participants (unwittingly) introduce. To do this, we moved any duplicate words from the SEEN set into the UNSEEN set before training each participant. So, if a particular subject had introduced polysemy by using the same word for more than one picture, only one of those word-picture pairs would be provided as training for the next subject.

The result of this small modification is dramatic. Although subjects find the task much harder and perfect transmission is not achieved, the learnability of the language nevertheless increases. However, the type of structure that emerges to make the language learnable is quite different. The strings in the language start to gain internal structure and in some cases clear compositionality emerges with aspects of the meaning being expressed as regular prefixes or suffixes. So, for example, in one of the languages that emerges, a black bouncing circle is named "winekuki" with the prefix "wi-" being largely consistently used to refer to black things, and the suffix "-kuki" being across the board to refer to anything bouncing. In this particular language, the shape is encoded by a complex set of semi-regularities governing the middle syllable and changes to the prefix (see Cornish, 2006, for the complete set of languages in the experiments).

It is important to reiterate that participants in this experiment are not deliberately constructing a structured system for encoding meanings (as they are in an experiment such as Selten & Warglien, 2007). They are attempting as best they can to give us back the language that they were exposed to, idiosyncrasies and all. In fact, some subjects reported that they were not even aware that they were being exposed during the test phase to pictures that they had not seen in training. In addition, the adaptation that occurs is not instantaneous, but gradual and cumulative. The increase in the learnability of the language tends to proceed by small amounts each generation.

This is truly an invisible hand process. The linguistic structure that emerges, which enables the subjects to accurately report the labels for pictures they have never seen, appears to be designed for that purpose, and yet there is no intentional designer. Just as in the computational and mathematical models, the mere fact that language must be passed through a transmission bottleneck causes it to adapt.

7. Conclusions

In this chapter, we have put forward the view that to explain the universal structural properties of language we need to look at language as a complex adaptive system – one in which biologically evolved innate biases on individual learning can be seen as challenges to which a culturally evolving language must adapt. Computational, mathematical and experimental models demonstrate that the process of linguistic evolution on a cultural time-scale is one that has significant explanatory power.

This growing body of work points to a number of conclusions of relevance to linguistics and the study of cultural evolution more broadly:

- biological evolution by natural selection is not the only explanation for adaptive structure in language;
- statistically significant cross-linguistic universals do not necessarily imply strong innate constraints;
- the burden of explaining the constraints on linguistic variation is lifted from a putative biologically evolved innate Universal Grammar;
- the structure of the human language faculty cannot be straightforwardly inferred from the observed structure of human language;

- the appearance of design in human behaviour, including language, does not necessarily require a designer if it is transmitted culturally.

In addition, we hope that we have made a case for attempting to model cultural evolution of language either in simulation or in an experimental setting. For too long explanatory formal frameworks for language structure have focussed on the individual and assumed that population effects are unimportant. It is now possible to move beyond these kinds of idealisations and explicitly examine what happens when populations of individuals interact. Work in this area is still in its infancy, but we believe it has the potential to improve our fundamental understanding of why language is structured the way it is.

References

Andersen, H. (1973). Abductive and deductive change. *Language*, 40:765–793.

Aunger, R., editor (2001). *Darwinizing Culture*. Oxford University Press.

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

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

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In Briscoe, T., editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, chapter 5. Cambridge University Press.

Bickerton, D. (1990). *Language and Species*. University of Chicago Press.

Bickerton, D. (2007). Language evolution: a brief guide for linguists. *Lingua*, 117(3):510–526.

Boyd, R. and Richerson, P. J. (1985). *Culture and the Evolutionary Process*. University of Chicago Press.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54.

Brighton, H. (2003). *Simplicity as a Driving Force in Linguistic Evolution*. PhD thesis, Theoretical and Applied Linguistics, The University of Edinburgh.

Brighton, H., Smith, K., and Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, 2:177–226.

Briscoe, E. J. (2000). Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language*, 76(2):245–296.

Cangelosi, A., Greco, A., and Harnad, S. (2002). Symbol grounding and the symbolic theft hypothesis. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*, pages 191–210. Springer Verlag.

Chomsky, N. (1975). *Reflections on Language*. Pantheon Books.

Christiansen, M. H. (1994). *Infinite Languages, Finite Minds: Connectionism, Learning and Linguistic Structure*. PhD thesis, University of Edinburgh, Scotland.

Cornish, H. (2006). Iterated learning with human subjects: an empirical framework for the emergence and cultural transmission of language. Master's thesis, University of Edinburgh.

Croft, W. (1990). *Typology and Universals*. Cambridge University Press.

Croft, W. (2000). Explaining Language Change: an evolutionary approach. Longman.

Deacon, T. W. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. W.W. Norton.

de Boer, B. (2001). *The Origins of Vowel Systems*. Oxford University Press.

Fay, N., Garrod, S., MacLeod, T., Lee, J., and Oberlander, J. (2004). Design, adaptation and convention: The emergence of higher order graphical representations. In *Proceedings of the 26th Annual Confernece of the Cognitive Science Society*, pages 411–416.

Francis, N. and Kucera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Hoghton Mifflin.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, 29:737–767.

Gomez, R. and Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4(5):178–186.

Griffiths, T. and Kalish, M. (2005). A bayesian view of language evolution by iterated learning. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society*.

Hawkins, J. A., editor (1988). *Explaining Language Universals*. Blackwell.

Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge University Press.

Healey, P., Garrod, S., Fay, N., Lee, J., and Oberlander, J. (2002). Interactional context in graphical communication. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pages 441–446.

Heine, B. and Kuteva, T. (2002). On the evolution of grammatical forms. In Wray, A., editor, *The Transition to Language*. Oxford University Press.

Hoekstra, T. and Kooij, J. (1988). The innateness hypothesis. In Hawkins, J. A., editor, *Explaining Language Universals*. Blackwell, Oxford.

Horner, V., Whiten, A., Flynn, E., and de Waal, F. B. M. (2006). Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proceedings of the National Academy of Sciences*, 103(37):13878–13883.

Hurford, J. (1990). Nativist and functional explanations in language acquisition. In Roca, I. M., editor, *Logical Issues in Language Acquisition*, pages 85–136. Foris, Dordrecht.

Hurford, J. (2000). Social transmission favours linguistic generalization. In Chris Knight, J. R. H. and Studdert-Kennedy, M., editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 324–352. Cambridge University Press, Cambridge.

Kalish, M. L., Griffiths, T. L., and Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review*, 14(2):288–294.

Keenan, E. and Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8:63–99.

Keller, R. (1994). *On Language Change: the invisible hand in language*. Routledge.

Kirby, S. (1994). Adaptive explanations for language universals: a model of hawkins' performance theory. *Sprachtypologie und Universalienforschung*, 47:186–210.

Kirby, S. (1997). Competing motivations and emergence: explaining implicational hierarchies. *Language Typology*, 1(1):5–32.

Kirby, S. (1999). *Function, Selection and Innateness: the Emergence of Language Universals*. Oxford University Press.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In Knight, C., editor, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 303–323. Cambridge University Press.

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

Kirby, S. (2002a). Learning, bottlenecks and the evolution of recursive syntax. In Briscoe, T., editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, chapter 6, pages 173–204. Cambridge University Press.

Kirby, S. (2002b). Natural language from artificial life. *Artificial Life*, 8(2):185–215.

Kirby, S. (2007). The evolution of meaning-space structure through iterated learning. In Lyon, C., Nehaniv, C., and Cangelosi, A., editors, *Emergence of Communication and Language*, pages 253–268. Springer Verlag.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. Forthcoming.

Kirby, S., Dowman, M., and Griffiths, T. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(12):5241–5245.

Kirby, S. and Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*, chapter 6, pages 121–148. Springer Verlag, London.

Kirby, S., Smith, K., and Brighton, H. (2004). From ug to universals: linguistic adaptation through iterated learning. *Studies in Language*, 28(3):587–607.

Mesoudi, A., Whiten, A., and Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Jorunal of Psychology*, 97:405–423.

Mesoudi, A., Whiten, A., and Laland, K. (2006). Towards a unified science of cultural evolution. *Behavioral and Brain Sciences*, 29:329–383.

Nowak, M. A., Komarova, N. L., and Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature*, 417:611–617.

Oudeyer, P.-Y. (2006). *Self-organization in the Evolution of Speech*. Oxford University Press.

Pinker, S. and Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13(4):707–784.

Selten, R. and Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences*, 104(18):7361–7366.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Strahan and Cadell.

Smith, A. (2005). The inferential transmission of language. *Adaptive Behavior*, 13(4):311–324.

Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science*, 14(1):65–84.

Smith, K., Kirby, S., and Brighton, H. (2003). Iterated learning: a framework for the emergence of language. *Artificial Life*, 9(4):371–386.

Sperber, D. (1996). Explaining Culture: A Naturalistic Approach. Blackwell.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Sciences*, 7(7):308–312.

Steels, L., Kaplan, F., McIntyre, A., and Looveren, J. V. (2002). Crucial factors in the origins of word-meaning. In Wray, A., editor, *The Transition to Language*. Oxford University Press.

Teal, T. and Taylor, C. (1999). Compression and adaptation. In Floreano, D., Nicoud, J.-D., and Mondada, F., editors, *Advances in Artificial Life*, pages 709–719. Springer.

Tonkes, B. (2001). *On the Origins of Linguistic Structure: Computational models of the evolution of language*. PhD thesis, University of Queensland, Australia.

Vogt, P. (2005). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence*, 167:206–242.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, 18(1):47–67.

Zuidema, W. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In Becker, S., Thrun, S., and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15 (Proceedings of NIPS'02)*, Cambridge, MA. MIT Press.

Language Learning ISSN 0023-8333

Complex Adaptive Systems and the Origins of Adaptive Structure: What Experiments Can Tell Us

Hannah Cornish
The University of Edinburgh

Monica Tamariz
The University of Edinburgh

Simon Kirby
The University of Edinburgh

Language is a product of both biological and cultural evolution. Clues to the origins of key structural properties of language can be found in the process of cultural transmission between learners. Recent experiments have shown that iterated learning by human participants in the laboratory transforms an initially unstructured artificial language into one containing regularities that make the system more learnable and stable over time. Here, we explore the process of iterated learning in more detail by demonstrating exactly how one type of structure—compositionality—emerges over the course of these experiments. We introduce a method to precisely quantify the increasing ability of a language to systematically encode associations between individual components of meanings and signals over time and we examine how the system as a whole evolves to avoid ambiguity in these associations and generate adaptive structure.

Hannah Cornish is funded by an ESRC postgraduate studentship. Monica Tamariz was funded by a Leverhulme Early Career Fellowship. Monica Tamariz and Simon Kirby are supported by ESRC Grant Number RES-062-23-1537 and by AHRC Grant Number AH/F017677/1. The work presented here benefited from discussion among the members of the University of Edinburgh's Language Evolution and Computation Research Unit, and constructive comments from two anonymous referees.

Correspondence concerning this article should be addressed to Simon Kirby, Language Evolution and Computation Research Unit, The University of Edinburgh, 3 Charles Street, Edinburgh, EH8 9AD, UK. Internet: simon@ling.ed.ac.uk

Introduction

The position paper in this issue (Beckner et al.) sets out a powerful motivating picture of language as a complex adaptive system. It presents us with a way of thinking not only about the dynamics of language as we know it now but also the emergence of language in the first place. More specifically, in this article, we suggest that the very fact that language persists through multiple repeated instances of usage can explain the origins of key structural properties that are universally present in language. Because of this, taking a complex adaptive systems perspective on language lifts the burden of explanation for these properties from a putative richly structured domain-specific substrate, of the sort assumed by much of generative linguistics (e.g., Chomsky, 1965). Ultimately, this alters our view of what biological evolution must provide in order to get language off the ground.

Much of the work over the past 20 years or so in modeling the evolution of language has taken this complex adaptive systems perspective (see, e.g., Brighton, Smith, & Kirby, 2005; Kirby, 2002b; Steels, 2003, for review). One particular strand of work has focused on the adaptation of language through a repeated cycle of learning and use within and across generations, where adaptation is taken to mean a process of optimization or *fitting* of the structure of language to the mechanisms of transmission (Kirby, 1999).[1]

A particular subset of models have looked extensively at the impact of repeated learning on the process of emergence. They investigate how a form of cultural evolution known as *iterated learning* affects the structure of language (e.g., Batali, 1998; Brighton, 2002; Griffiths & Kalish, 2007; Hurford, 2000; Kirby, 1999; Kirby, Dowman, & Griffiths, 2007; Kirby & Hurford, 2002; A. Smith, 2005; K. Smith, 2002; Vogt, 2005; Zuidema, 2003). In these models, each agent (i.e., simulated individual) must acquire a set of (initially random) mappings between meanings and signals by observing the behavior of agents in the previous generation. Once this mapping is acquired, the learner becomes a teacher, and the process repeats. Crucially there is a bottleneck in the transmission process that puts pressure on the system to be generalizable (Deacon, 1997). This bottleneck models the data-sparsity present in real language acquisition and is typically enforced in the simulations by the learner only being exposed to signals for a subset of the total meanings during training.

Overall, two consistent conclusions have been drawn from this computational research: Over time, iterated learning ensures languages evolve to (a) become easier to learn and (b) become more structured. These two facts are not unrelated: One of the ways in which a language can evolve to become more learnable is by becoming structured. This is because there are only two ways to survive the transmission bottleneck: be heard (and remembered) by the next generation or be easily inferable from what is heard. This latter solution can only occur when there is some kind of regularity to be exploited in the system. The exact form this regularity takes can vary, which is something we explore later.

The regularity that emerges gradually in the computational simulations justifies our use of the term "adaptive" in this case. This is because the kinds of linguistic structures that evolve show the hallmarks of *apparent design*. For example, in some models (e.g., Batali, 2002; Kirby, 2002a), recursive compositional syntax evolves that clearly enables the simulated agents to successfully convey meanings in an open-ended way. This kind of adaptive structure in language might lead researchers to conclude that it must reflect innate constraints that are the result of biological evolution by natural selection (e.g., Pinker & Bloom, 1990). However, this conclusion is not justified. In most of these models, there is no biological evolution. Indeed, individuals are essentially clones throughout. Rather, the adaptation arises purely from the iterated learning process itself. Language transmission is a complex adaptive system.

Recently, we developed a method for studying this process of adaptive evolution in the laboratory, extending experimental studies of iterated learning in the nonlinguistic domain by Griffiths and Kalish (2007) and Kalish, Griffiths, and Lewandowsky (2007). By combining two experimental techniques—artificial language learning (e.g., Esper, 1925, 1966; Fitch & Hauser, 2004; Gómez & Gerkin, 2000; Saffran, Aslin, & Newport, 1996) and diffusion chains (e.g., Bangerter, 2000; Bartlett, 1932; Horner, Whiten, Flynn, & de Waal, 2006; Mesoudi, Whiten, & Dunbar, 2006; Whiten, Horner, & de Waal, 2005)—we were able to track the evolution of a miniature language over "generations" of experimental participants from an initially random, unstructured state, to one showing clear evidence of adaptive structure (Kirby, Cornish, & Smith, 2008).[2] In this article, we provide a new analysis of the results of this study to examine in more detail the way structure emerges as a result of competition between linguistic variants.

Human Iterated Learning: An Overview

Before we move onto the details of the studies, it is necessary to familiarize ourselves with the general methodology and key parameters of the experiments that follow. A participant is trained on an "alien" language consisting of a set of meanings (usually presented as pictures) paired with signals (a string of letters, or possibly sounds) drawn from a finite set. After being trained on some proportion of these meanings, the participant is then presented with a series of meanings without signals and asked to provide the correct description in the alien language. These meanings and signals are recorded and become the new set of training pairs for the next participant, who forms the next "generation" of the chain. This procedure is repeated until the chain is complete (i.e., until the desired number of generations has been reached).

Participants involved in the study are only asked to learn the language as best they can: They are not told anything about the iterated nature of the study or that their responses will be given to future participants. During each training round, participants are shown a picture drawn at random from the set of meanings, and below it, a string of letters that they are told represents how the alien would describe that picture in its own language. Training occurs via a computer, and each exposure is timed to ensure no training item (meaning-signal pair) is seen for longer than any other and continues until all training items have been seen. During the final test, the participant is shown each picture in the language once, one after another, and asked to type in the missing descriptions. These responses are then randomly sampled from to generate the new training items for the next generation.

Clearly, this experimental setup represents a highly simplified idealization of the real process of linguistic transmission. In particular, the population model is the simplest that we could construct (in line with the other diffusion chain experiments mentioned previously). Three parameters characterize possible population models: direction of transmission (vertical or horizontal), the size of the population, and who learns from whom (network structure). For the rest of this article we focus on just one scenario: vertical transmission, involving 10 people, with each person learning from just 1 other person. However, it is important to remember that there are many other scenarios that could be explored within this framework.

Learnability, Expressivity, and Adaptation

As stated in the introduction, the main finding to have emerged over the past decade or so of research into this area is that languages themselves adapt to be better learnable and transmissible by us over time (see, e.g., Christiansen & Chater, 2008, for a review). However, it should be recognized that this pressure toward greater learnability must be tempered somewhat in order for structure to emerge. The reason for this is simple: The most easily learnable language might be one in which there is one word for everything (or possibly, no words at all). It is only when we also have a pressure for expressivity, for meanings to actually be distinguished from one another, that we are likely to see the emergence of structure.

The first application of this new experimental methodology set about investigating this tension between expressivity and learnability (Kirby et al., 2008). In this study, the meaning space consisted of 27 pictures showing a scene that varied along three features and three values: color of object (blue, black, red), shape of object (circle, triangle, square), and a dotted line indicating the movement of object (bouncing, spiralling, moving horizontally). Two different experimental conditions were explored, with four chains of 10 people in each. In one condition there was a "hidden" pressure for each of the meanings in the meaning space to be expressed uniquely: Participants' input was filtered in such a way as to ensure they never perceived different meanings with the same signal. In the other, there was no such pressure. Participants could not be aware of the experimental condition in which they were included.

The chains in each condition both began with random initial languages, and a transmission bottleneck was imposed by exposing each generation with just half (14) of the meaning-signal pairs during training (the particular meanings that they would be exposed to were chosen randomly each generation). Example (1) shows a sample of the initial randomly generated language in one of the chains to illustrate what is meant by the claim that they are unstructured with respect to their meanings.[3] In spite of the fact that these meanings in the world are similar (triangles of every color that either move horizontally or in a spiral), the signals used to describe them are all idiosyncratic, with no consistently repeating subparts.

(1)

a.	kapihu "black triangle horizontal"	b.	luki "black triangle spiral"
c.	humo "blue triangle horizontal"	d.	namola "blue triangle spiral"
e.	lahupiki "red triangle horizontal"	f.	lumoka "red triangle spiral"

After training, participants were tested on all 27 meanings, and it is from this output set that the new training set is sampled for the participant in the next generation.

The main findings can be summarized as follows (see Kirby et al., 2008, for more details). First, by looking at the learning errors made between adjacent generations, it was shown that the languages in both conditions were being acquired significantly more faithfully toward the end of the chains than they were at the beginning. Second, this increase in learnability over time occurred as a result of the languages becoming more structured over time.

What is interesting about this last fact, however, is that the way in which the languages were structured differed markedly between the two experimental conditions. In the first condition, for which there was no filtering of the participants' input, systems emerged that were characterized by *underspecification*. This involved a reduction in the total number of distinct signals, introducing ambiguity with respect to the meanings. However, this ambiguity was not complete, as it did not affect all meaning dimensions. In one chain for instance, a system emerged [of which a sample is reproduced as Example (2)] whereby everything that moved horizontally was called *tuge*, everything that moved in a spiral was named *poi*, and there was a three-way distinction of bouncing items dependent on shape: for bouncing squares, *tupim* for bouncing triangles, *tupin*, and for bouncing circles, *miniku*. This system proved to be highly adaptive in the sense that, once it emerged, it was stable and faithfully acquired by subsequent generations without error.[4]

(2) a. tuge
"black triangle horizontal"

b. poi
"black triangle spiral"

c. tuge
"blue triangle horizontal"

d. poi
"blue triangle spiral"

e. tuge
"red triangle horizontal"

f. poi
"red triangle spiral"

As Kirby et al. (2008) pointed out, underspecification is not an unusual feature of human languages, but taken to extremes, it would lead to an inexpressive and communicatively disfunctional language (albeit one that would be easy to learn). The second experimental condition, whereby items were removed from a participant's input if they should lead to the same string being assigned to more than one meaning, was designed to introduce a hidden pressure against underspecification. With this modification in place, the systems that emerged appear much closer to what we might expect a communicatively useful system to look like. These systems were characterized by *compositionality*, whereby the meaning of a given string could be inferred by the meaning of subparts of that string (morphemes) and the way they are put together. Example (3) again shows a sample of this.[5]

(3) a. nekeki
"black triangle horizontal"

b. nekipilu
"black triangle spiral"

c. lakeki
"blue triangle horizontal"

d. lakipilu
"blue triangle spiral"

e. raheki
"red triangle horizontal"

f. rahopilu
"red triangle spiral"

These results are very exciting, as they experimentally verify the main findings to have emerged from computational models of iterated learning for the first time: that languages adapt purely by virtue of transmission through iterated learning. Moreover, the kind of adaptation is determined, in part, by constraints placed on the transmission of the languages about which participants could not be aware. However, although it has been shown that the languages in these experiments *do* adapt, it has not yet been established *how* they adapt. It is to this question that we now turn.

The Evolution of Signals During Iterated Learning

In this subsection we will focus on the utterances, leaving aside the meanings for the moment, and construct phylogenies demonstrating the evolution of linguistic forms over iterations. We used one of the languages [part of which was reproduced in Example (2)], taken from Kirby et al. (2008) to construct the coalescent tree shown in Figure 1. These trees are a standard way to represent phylogenetic descent in evolutionary biology (Barton, 2007; Hein, Schierup, & Wiuf, 2005), although here we have amended them to also include frequency information in brackets. Bold lines show perfect replication of an utterance, whereas other lines show possible relationships of descent between utterances across generations.

As we can see in Figure 1, the number of different utterances decreases over time as we start to observe perfect replication of select utterances, along with a general tendency for utterances to become shorter. In the early history of this language, the process of transmission is principally one of generating new recombinations of signal substrings. We observe only one instance of replication of a whole utterance but many replications of parts of the utterances, such as unigrams or bigrams, and even larger *n*-grams. For example, the introduction of the form miniku in generation 2 could be the result of a blend between miniki and miweniku.[6] There is still much variation in the language at this point. In the final generations, however, the frequencies of the few remaining units stabilize around multiples of 3, suggesting adaptation to a meaning space containing three dimensions.

In the case of the language in Figure 1, given the nondecomposable utterances that survived into the final stable system, it was appropriate to analyze

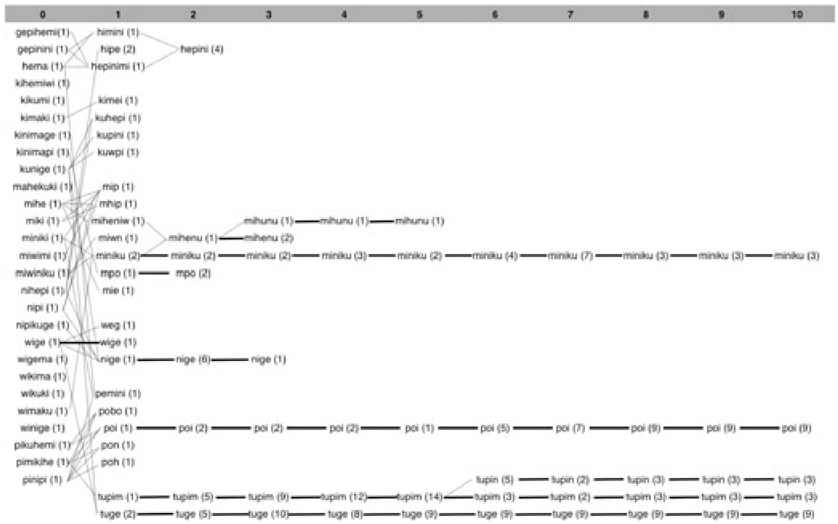


Figure 1 Coalescent tree showing lineages of signals for all 27 items over generations of one of the languages obtained by Kirby et al. (2008) exhibiting systematic underspecification. Columns correspond to generations; horizontal bold lines indicate the perfect replication of the whole signal; all other lines indicate some of the possible relationships of descent between signals that share some features. Numbers shown in brackets indicate the frequency with which variants were produced at each generation. The number of variant types decreases over time, although the number of tokens remains fixed at 27 throughout. Among these surviving variants there are clear relationships of descent, sometimes with modification. The frequency information is suggestive of the fact that signal variants may be adapting to express a meaning space composed of multiples of 3.

replication at the level of the whole utterance. However, in a compositional system, the meaning of a complex utterance is a function of the meanings of the elements of the utterance and the way they are arranged. The tree in Figure 1 illustrates adaptation of the whole signals to the structure of the meaning space; in a compositional language, we expect the same phenomena to occur but this time at the level of signal elements. We will now quantify compositionality using a different language (part of which is shown in Example (3)] from Kirby et al. (2008).

First, we need to segment the signals into element units. To do this, we first examined the language of the final participant in the chain to find the most parsimonious segmentation of the strings into elements that corresponded to aspects of the meanings (e.g., "the signal endings reliably encode motion" or "signal-initial 'la' consistently encodes colour blue"). This resulted in each

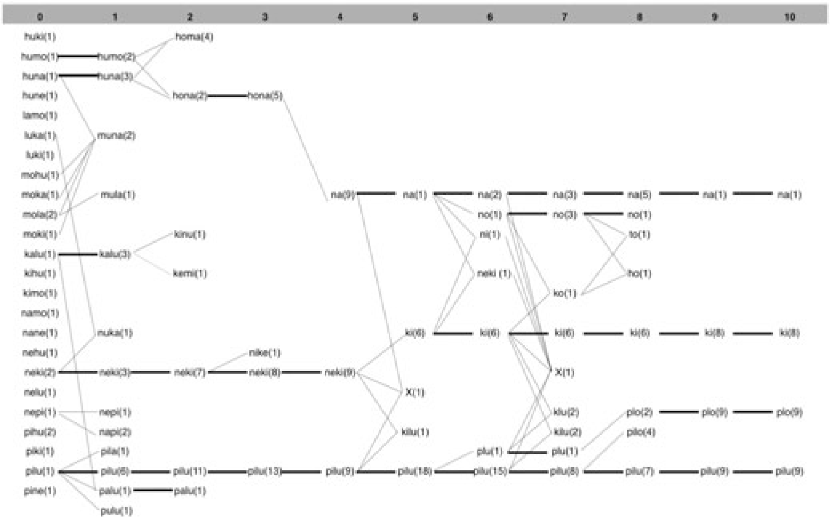


Figure 2 Coalescent tree showing lineages of signal ending variants for all 27 items over 10 generations of one of the languages obtained by Kirby et al. (2008), referred to in (3). Numbers shown in brackets indicate the frequency with which variants were produced at each generation. The number of variant types decreases over time, although the number of tokens remains fixed at 27 throughout. Among these surviving variants there are clear relationships of descent, sometimes with modification. The frequency information is suggestive of the fact that signal variants may be adapting to express a meaning space consisting of three meaning elements (see generations 4, 9, and 10).

string being divided into three substrings, and this segmentation pattern was carried back to all previous generations in order to allow for a consistent analysis. Figure 2 shows the coalescent tree for the word-final signal element (although similar trees can be constructed for both initial and middle positions also).

As earlier, we observe a marked reduction in the number of variants over time, as just a few become selected to be reused more often. Furthermore, we can see that the variants that appear at each generation are not random; we can trace the genealogy of the surviving variants back in time. Even over this minute timescale, many of the changes observed appear to follow paths that are well attested in cases of natural language change, such as single segment replacements (nepi → napi; pilu → pilo), reductions (hona → na, neki → ki, pilu → plu), metathesis (neki → nike), and blends (humo & huna → homa & hona; na & ki → neki).

It is significant to notice that at generation 4 we have three variants (na, neki, pilu), each with a frequency of 9 and that for the final two generations, this pattern repeats (now for variants ki, plo, pilu) broken only by a single instance of na. This, again, suggests that these lineages are adapting to a three-element meaning space. Obviously, we know that this is indeed the case; the interesting thing is that the signals alone suggest it. In the next subsection, we show how we can precisely quantify regularities in the mappings between signal and meaning elements in order to objectively confirm this.

Quantifying the Emergence of Compositionality

We now have an analysis of all the languages in terms of the following: *signal segments*—in this case, the word beginning, middle, or end; *signal segment variants*—actual tokens residing in a segment position, such as pilu or ki. Similarly, we can define the following: *meaning elements*—aspects of meaning, such as motion, shape, and color; *meaning element variants*—actual instances of a meaning element, for instance, “blue,” or “circle,” or “bounce.”

Kirby et al. (2008) quantified the emergence of structure using a pairwise distance correlation (Shillcock, Kirby, McDonald, & Brew, 2001). This measures the extent to which similar meanings are expressed using similar forms—or more precisely, whether there is a correlation between the structure of the meaning and signal spaces. Although this is valuable in showing that structure emerges, it does not allow us to track the evolution of the compositional structure of the languages directly: As a measurement, the pairwise distance correlation is very general and cannot distinguish between compositionality and other kinds of structures (such as underspecification). Here, we apply a new method of analysis to one of the chains[7] reported in Kirby et al. (2008) to tackle this problem. We use *RegMap* (Tamariz & Smith, 2008), an information-theoretic metric that combines the conditional entropy of meanings given signals and of signals given meanings and normalizes the result to make it comparable across systems of different sizes. Informally, what *RegMap* (short for regularity of the mappings) does is return the degree of confidence that a signal element consistently predicts a meaning element (for instance, the degree to which we can be sure that the beginning of the signal encodes color).

More formally, $H(X \mid Y)$, the conditional entropy, is the Shannon entropy (Shannon, 1948) but replacing $p(x)$ with $p(x \mid y)$. The *RegMap* for a meaning element (M) and a signal segment (S) is given by

$$\mathrm{Reg}Map = \sqrt{\left(1 - \frac{H(S \mid M)}{\log(n_s)}\right) \times \left(1 - \frac{H(M \mid S)}{\log(n_m)}\right)}. \quad (1)$$

$H(S\,|\,M)$ is the conditional entropy of the signal segment given the meaning feature, or the uncertainty about the meaning when we know the segment. This relates to comprehension. For example, for shape and first signal segment, $H(S\,|\,M)$ quantifies how uncertain we are on average about what shape an object is if we hear the first segment of its corresponding signal. $H(M\,|\,S)$ is the conditional entropy of the meaning feature given the signal segment, or the uncertainty about the segment when we know the meaning. This relates to production. Still, in the case of shape and first signal segment, $H(M\,|\,S)$ quantifies how uncertain we are, on average, about what first segment to produce if we know the shape of an object. The logs of n_m and n_s normalize the values between 0 and 1; n_m is the number of different meaning values (e.g., triangle, circle, square for shape); n_s is the number of different segment variants in the relevant segment position. Subtracting the conditional entropies from Equation 1 returns levels of confidence instead of uncertainty.

Figure 3 shows the *RegMap* values for all combinations of signal and meaning elements both with and without a bottleneck for the 10 generations. The "input" data shown in Figure 3 (upper) reflects the extent to which signals predict meanings in the subset of the language (taken from the previous generation) that was actually transmitted to the current generation, after the bottleneck was applied. The "output" data shown in Figure 3 (lower) is obtained from the complete languages that participants actually produced at a given generation, before the bottleneck was applied. The significance of the obtained *RegMaps* was established with a Monte Carlo analysis involving 10,000 randomizations of the correspondences between meanings and signals and are shown as boxplots.

Focusing first on the bottom graphs of Figure 3, obtained from the participants' output languages, we see that, starting from values indistinguishable from random at generation 1, *RegMap* becomes massively increased to highly statistically significant levels; specifically, by the third generation, motion is consistently encoded by the final signal segment; by the fourth generation, color is encoded by the initial segment, and by the ninth generation, shape is encoded by the middle segment (all $p < .001$).

Second, a comparison of the input (upper) and output (lower) results in Figure 3 reveals the effect of the bottleneck. The *RegMap* values are, in the majority of cases, amplified by the bottleneck (the absolute value of *RegMap* increases). Moreover, the lower the input *RegMap*, the more likely it is to be amplified by the bottleneck. How is this happening? The answer is potentially counterintuitive; randomly occurring patterns are more likely to be perceived the smaller the system is. At least in the early generations, a subset drawn from a language is more likely to accidentally contain more regular patterns than the

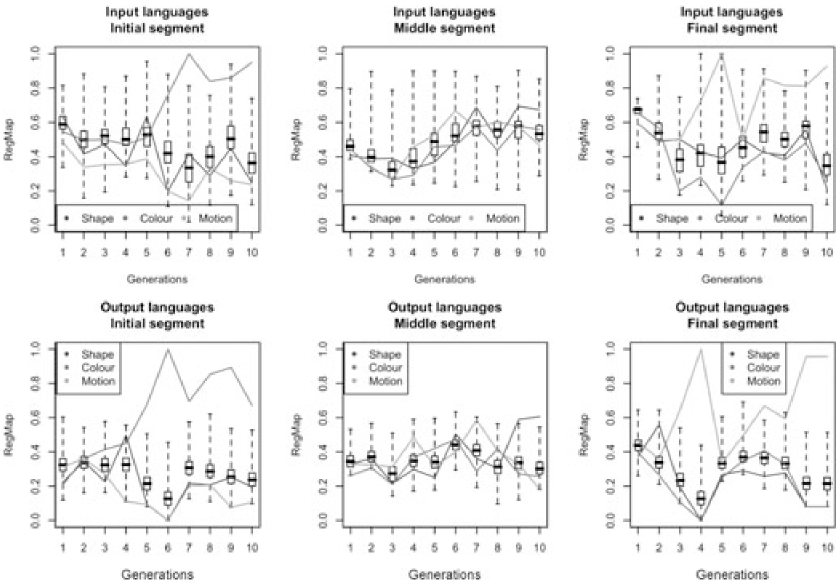


Figure 3 Regularity of the associations between signal and meaning elements, measured as *RegMap*, changes over time in the direction of maximizing compositionality, whereby signal elements are consistently associated with distinct meaning elements. The continuous colored lines represent *RegMap* values obtained with all nine segment-meaning feature pairs in the 10 generations of a language family from the study by Kirby et al. (2008), referred to in Example (3). The boxplots show the distributions of values obtained with 10,000 randomized languages. The upper graphs show *RegMap* values from the subset of language (taken from the previous generation) that was actually transmitted to the current generation, after the "bottleneck" was applied. The lower graphs show *RegMap* values obtained from the complete languages that participants actually produced at a given generation, before the bottleneck was applied.

entire language. Implicit in this, and by the same token, a given subset will also tend to contain less counterevidence against such patterns. This explains why we observe such a dramatic difference between the ranges shown in the boxplots in the upper and lower graphs in Figure 3. The large range of *RegMap* values in the input languages directly reflects the fact that participants are sensitive to this reduced number of observations when they are inferring the mappings between meanings and signals. Together, this accounts for the structure-generating effect of the bottleneck on language: The input to each generation is only a fraction of (and therefore tends to be more systematic than) the total output of the previous generation.

Third, the graphs show cases of competition between meanings all vying to be expressed by the same signal element. For example, motion and shape are both equally encoded in the final signal segment in the input to generation 3, but generation 3 resolves this conflict by ignoring one of the associations (shape) and amplifying the other (motion) to a significance level of $p < .01$. Conversely, we also see cases of competition between signals vying to express the same meaning: In the input of generation 5, color is equally encoded in the initial and middle signal segments (similar absolute values and levels of significance); in this case, the conflict is resolved by massively amplifying the association with the initial segment to a significance level of $p < .001$ and reducing the association with the middle one. These processes are adequately explained by the standard evolutionary mechanisms of variation, replication, and selection applied to the mappings between signals and meanings elements. Selection, in this case, can be hypothesized to be guided by perceptual and attentional biases such as higher salience of certain signal and meaning elements over others. Unfortunately, a detailed discussion of these biases is outside the scope of the present work.

Summary

Kirby et al. (2008) found that the languages that emerge through a repeated cycle of learning and production in a laboratory setting show evidence of adaptation to the bottleneck placed on their transmission. Making even minor changes to the way in which language is culturally transmitted can produce radically different types of structures. Given only a bottleneck on transmission preventing a proportion of the language from being seen by the next generation, language can adapt in such a way that ensures that it is stably transmitted to future generations. However, this occurs at the expense of being able to uniquely refer to every meaning. When they introduced the additional pressure of having to use a unique signal for each meaning, the language once again adapted to cope with these new transmission constraints, this time by becoming compositional. Having a compositional system ensures that both signals and meanings survive the bottleneck.

Because the participants could not know which condition they were in, it is impossible that the resulting languages were intentionally designed as adaptive solutions to the transmission bottleneck. Rather, the best explanation for the result is that in these experiments, just as in the computational models, linguistic adaptation is an inevitable consequence of the transmission of linguistic variants under particular constraints on replication. The result is apparent design, but without an intentional designer.

Whereas Kirby et al. (2008) analyzed their results at the level of whole signals and whole meanings, in this subsection we have developed new techniques to analyze the same results in terms of the component parts of linguistic signals. An analysis of how signal variants and their frequencies change over time showed relationships of descent with modification among them. It also suggested that signal variants are adapting to the structure of the meaning space. This intuition was verified by the application of *RegMap*, a tool designed to objectively measure compositionality. Using this method, we showed that individual signal elements come to encode individual meaning elements, whereas the whole system evolves to avoid ambiguity (i.e., more than one meaning being encoded in the same signal element or vice versa). Moreover, we were able to more precisely describe the role of the bottleneck in bringing about compositionality: The smaller subsets sampled as inputs to the next generation may locally contain more systematicity than the entire language. Iterating this learning process using these smaller samples therefore provides a platform that allows systematic patterns to be noticed, remembered, and replicated preferentially, thereby allowing them to gradually accumulate in the language as a whole.

It seems clear from all of this that, first, cultural transmission alone is capable of explaining the emergence of languages that exhibit that appearance of design and, second, experimental studies of the iterated learning of artificial languages are a potentially useful methodological tool for those interested in studying cultural evolution.

Conclusion

This article has extended previous work on iterated language learning experiments by showing, using data obtained from an earlier study, exactly how compositional structure emerges over time as a result of cultural transmission. Using a recently developed analytical technique that calculates the regularity of mapping between signal and meaning elements (Tamariz & Smith, 2008), we were able to precisely quantify changes in the language's ability to systematically encode such associations between meaning and signal components. From this we were able to explain the amplification effect the bottleneck seems to have on systematicity in language, arguing that the sampling of smaller subsets of the language for training input to the next generation tends to make weaker patterns that are not visible at the level of the entire language appear stronger locally.

One obvious criticism of the experimental work described here is that it necessarily involves participants who already speak a language. As such, can it tell us anything about the original evolution of language, as we are claiming? The sceptical position might be that we are simply seeing the evolution of structure that reflects the native language of the participants as opposed to any adaptive logic of the iterated learning process itself. This criticism faces a number of problems, however. Most importantly, the experimental results are backed up by the computational simulations and mathematical models surveyed in the introduction. In these models we can be sure that there is no influence of prior language, as the models have none initially. Furthermore, the structure that arises depends on aspects of the transmission bottleneck that are hidden from our participants (given our two experimental conditions) and the particular properties of the language appear more dramatically shaped by these than any similarity to the language of the participants. The most parsimonious explanation, then, is that we are seeing adaptation to the transmission bottleneck rather than an emerging simple first language influence. However, a more subtle point can be made: We fully expect that language evolution through iterated learning will involve adaptation to all aspects of the transmission bottleneck, and this will include the biases of language learners. In our experiment, participants bring to bear a mixture of biologically basic biases and those that arise from their acquired cultural heritage. We can see no principled way to separate these out. This means that our experiments should not be taken to be a "discovery procedure" for uncovering our evolutionary ancient learning biases but rather as a tool for understanding the fundamental adaptive dynamics of the cultural transmission of language by iterated learning.

We started this article by noting that a complex adaptive systems perspective shifts the burden of explanation away from a richly structured domain-specific innate substrate for language in our species. Although we have talked a great deal about linguistic structure as an adaptation, this is adaptation by the language itself rather than biological evolution of the faculty of language. The relevant explanatory mechanisms relate to *cultural* as opposed to *natural* selection. However, of course, this does not mean that biology is irrelevant to the evolution of language.

Rather than seeking evolutionary explanations for innate constraints that determine language structure, the work presented in this article strongly suggests a different approach. The iterated learning models on which we base our experiments start with agents who can (a) learn complex signals and (b) infer complex meanings. Humans belong to an unusual set of species, called the "vocal learners" (Jarvis, 2004), that can learn sequential signals (others include most notably the songbirds). We are also unusually adept in inferring intentionality (Tomasello, Carpenter, Call, Behne, & Moll, 2005). By taking into account the power of language as a complex adaptive system to generate structure itself, future work on the biological evolution of language in our species should focus on how we came to have these two crucial preadaptations for language. Without the combination of vocal learning and meaning inference, iterated learning of the sort we are studying would not be possible at all (Okanoya, 2002). Once they are in place, on the other hand, the emergence of structure is inevitable.

Revised version accepted 11 June 2009

Notes

1 Underlying this work is a typically unstated assumption that modern languages are already optimized for transmission (i.e., all extant languages are both learnable by children and meet the expressive needs of their users); thus, further change is driven not so much by inherent properties of linguistic variants but rather sociolinguistic factors (e.g., Croft, 2000). However, when looking at the origins of language, we necessarily need to consider a different state of affairs, one in which language has not yet reached equilibrium and the *inherent* structural properties of linguistic variants are relevant. A related point is the likelihood that intergenerational transmission is less important in ongoing language change than it is in language emergence. Where social status, for example, is the primary driving force behind selection of variants, the impact of learners' innovations is likely to be lower than where those innovations actually make language transmissible at all.

2 There are other experimental approaches to the origins of language, such as Galantucci (2005) and Selten and Warglien (2007), but note that these rely on participants intentionally and consciously designing a communicative system. Our interest is in whether the adaptive structure of language can arise without intentional design.

3 The glosses here are given as English words; recall that in the experiment, visual stimuli were used. This example is taken from Chain 3 in Experiment 2 in the study by Kirby et al. (2008).

4 This is not a trivial result considering the rather narrow bottleneck applied during training meant that each generation was being trained on a (different) random subset of half of the total language.

5 Taken from generation 9, chain 3, experiment 2 in the study by Kirby et al. (2008). Note that whereas color and motion are consistently expressed (ne for black, la for blue, ra for red, ki for horizontal, and pilu for spiral), shape is more haphazardly encoded (ke when blue/black and horizontal, ki when blue/black and spiral, he when red and horizontal, and ho when red and spiral).

6 It is perhaps interesting to note that the investigation of this type of phenomenon, historically referred to as analogical change, was what prompted the very first application of this methodology by Esper in 1925.

7 Specifically, we examine chain 3 in experiment 2, but similar results can be obtained wherever compositionality clearly emerges.

References

Bangerter, A. (2000). Transformation between scientific and social representations of conception: The method of serial reproduction. *British Journal of Social Psychology, 39*, 521–535.

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Barton, N. H. (2007). *Evolution*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 405–426). Cambridge: Cambridge University Press.

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 111–172). Cambridge: Cambridge University Press.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life, 8*(1), 25–54.

Brighton, H., Smith, K., & Kirby, S. (2005). Language as evolutionary system. *Physics of Life Reviews, 2*, 177–226.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Science, 31*(5), 489–509.

Croft, W. (2000). *Explaining language change: An evolutionary approach.* London: Longman.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: W.W. Norton.

Esper, E. A. (1925). *A technique for the experimental investigation of associative interference in artificial linguistic material*. Philadelphia: Linguistic Society of America.

Esper, E. A. (1966). Social transmission of an artificial language. *Language, 42*(3), 575–580.

Fitch, T., & Hauser, M. (2004). Computational constraints on syntactic processing in a non-human primate. *Science,* 303, 377–380.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science, 29*, 737–767.

Gómez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*(5), 178–186.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science, 31*(3), 441–480.

Hein, J., Schierup, M. H., & Wiuf, C. (2005). *Gene genealogies, variation and evolution: A primer in coalescent theory*. Oxford: Oxford University Press.

Horner, V., Whiten, A., Flynn, E., & de Waal, F. B. M. (2006). Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *PNAS, 103*(37), 13,878–13,883.

Hurford, J. (2000). Social transmission favours linguistic generalization. In J. R. H. Chris Knight & M. Studdert-Kennedy (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 324–352). Cambridge: Cambridge University Press.

Jarvis, E. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences, 1016*, 749–777.

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review, 14*(2), 288–294.

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

Kirby, S. (2002a). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 173–204). Cambridge: Cambridge University Press.

Kirby, S. (2002b). Natural language from artificial life. *Artificial Life, 8*(2), 185–215.

Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–148). London: Springer-Verlag.

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

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 10,681–10,686.

Mesoudi, A., Whiten, A., & Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology, 97*, 405–423.

Okanoya, K. (2002). Sexual display as a syntactical vehicle: The evolution of syntax in birdsong and human language through sexual selection. In A. Wray (Ed.), *The transition to language* (pp. 46–63). Oxford: Oxford University Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13*(4), 707–784.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month olds. *Science, 274*, 1926–1928.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental co-ordination game. *PNAS, 104*(18), 7361–7366.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal, 27*, 379–423, 623–656.

Shillcock, R., Kirby, S., McDonald, S., & Brew, C. (2001, August). Filled pauses and their status in the mental lexicon. In *Proceedings of the 2001 conference on disfluency in spontaneous speech*, Edinburgh, UK.

Smith, A. (2005). The inferential transmission of language. *Adaptive Behavior, 13*(4), 311–324.

Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science, 14*(1), 65–84.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Science, 7*(7), 308–312.

Tamariz, M., & Smith, A. D. M. (2008). Regularity in mappings between signals and meanings. In A. D. M. Smith, K. Smith, & R. Ferrer i Cancho (Eds.), *The evolution of language: Proceedings of the 7th international conference* (*EVOLANG7*) (pp. 315–322). Singapore: World Scientific.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675–691.

Whiten, A., Horner, V., & de Waal, M. (2005). Conformity to cultural norms of tool use in chimpanzees. *Nature, 437*, 737–740.

Vogt, P. (2005). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence, 167*, 206–242.

Zuidema, W. (2003). How the poverty of the stimulus argument solves the poverty of the stimulus argument. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (*Proceedings of NIPS'02*) (pp. 51–58). Cambridge, MA: MIT Press.

Investigating how cultural transmission leads to the appearance of design without a designer in human communication systems

Hannah Cornish
University of Edinburgh

Recent work on the emergence and evolution of human communication has focused on getting novel systems to evolve from scratch in the laboratory. Many of these studies have adopted an interactive construction approach, whereby pairs of participants repeatedly interact with one another to gradually develop their own communication system whilst engaged in some shared task. This paper describes four recent studies that take a different approach, showing how adaptive structure can emerge purely as a result of cultural transmission through single chains of learners. By removing elements of interactive communication and focusing only on the way in which language is repeatedly acquired by learners, we hope to gain a better understanding of how useful structural properties of language could have emerged without being intentionally designed or innovated.

Keywords: iterated learning, cultural evolution, language emergence, linguistic transmission

1. Introduction

There has recently been renewed interest in studying the emergence and evolution of human communication systems experimentally (e.g. Galantucci, 2005; Garrod et al., 2007; Healey et al., 2007; Scott-Phillips et al., 2009; Selten & Warglien, 2007; Kirby et al., 2008a). These studies differ from the many experiments investigating human communication that went before (e.g. Garrod & Anderson, 1987; Garrod & Doherty, 1994; Christiansen, 2000; Pickering & Garrod, 2004; Hudson-Kam & Newport, 2005; Wonnacott & Newport, 2005) by the emphasis placed on exploring the emergence of *novel* systems.[1] In other words, these experiments do not start with a system (either natural or designed by the experimenter) in place initially, but lets one evolve over the course of the experiment. This provides us with a direct route into understanding how such systems become established (Galantucci, 2005).

Interaction Studies 11:1 (2010), **112–137**. DOI 10.1075/is.11.1.02cor
ISSN 1572–0373 / E-ISSN 1572–0381

Prior to this, researchers intent on understanding the complicated processes giving rise to systems of human communication were to some extent limited by a lack of data. The simple fact is that natural languages do not emerge every day, and neither do they emerge overnight. This left two options: (1) exploring the few documented cases of large-scale natural language emergence, involving sign-languages (Senghas & Coppola, 2001; Sandler et al., 2005), and home-sign (Goldin-Meadow & Mylander, 1983) that are available, or (2) exploring the phenomenon more abstractly in computational simulations (Batali, 1998; de Boer, 2001; Kirby, 2001; Steels, 2003; Brighton et al., 2005; Oudeyer, 2005; Vogt, 2005) or mathematical models (Griffiths & Kalish, 2007; Kirby et al., 2007).

It is clear that an experimental approach offers certain advantages over studying these phenomena indirectly via the use of computational/mathematical models, or via naturalistic observation (such as greater experimental manipulation, control, and replicability of results, etc.). Most of these newer experiments looking at the emergence of novel systems share the property of revolving around some kind of communication game. Participants (typically dyads) are given some shared goal or joint task that requires them to co-ordinate their actions in some way. The only way in which to do this is to interactively construct a communication system together, using whatever medium is provided.

For instance, in Selten & Warglien (2007), pairs of participants are given a repertoire of available symbols, each with different sending costs, and instructed to converge upon a set of economical signals to identify different pictures. In Galantucci (2005), pairs of participants must coordinate their actions in a 2D game-world by communicating with one another using a novel graphical medium, which prevents the use of common symbols or pictorial representations, forcing them to develop a new system of their own. In Healey et al. (2007), pairs of participants (and later on, interacting groups) collaborate together using a virtual whiteboard, drawing images to identify different pieces of music. Similarly, Garrod et al. (2007) encourage participants to depict various concepts (such as commonly known people, places, objects, and more abstract concepts such as 'poverty') using images in such a way that a fellow participant could identify them. In a slightly different twist, Scott-Phillips et al. (2009) have an experimental set-up in which they do not even provide a dedicated channel for communication to take place in: given a task which requires two players to coordinate their actions, the only solution is to create one by using their movements in the game environment as signals.

The fact that convergence does not come easily to participants in these experiments (most fail to agree on a system, and fewer still go on to develop one with structure) highlights the fact that the underlying processes responsible are not trivial. This is perhaps surprising given that we assume participants could easily invent a workable system on their own. In fact, Scott-Phillips et al. (2009) find that reported reasons for failure often centre around an inability to convey a system to their partner rather than an inability to individually construct one in the first place. Conversely, Selten & Warglien (2007) showed that the chances of developing a successful system are massively increased when one player finds a way to take control and impose their invented system upon the other. This raises the interesting question of what kind of design process we think is responsible for the emergence of structure in natural language – is it one which is wholly reliant on the ingenuity and design skills of its users, or is there some other force at work?

This paper argues that although humans are extremely adept at constructing novel communication systems, many linguistic changes are not 'designed' by individuals in that manner. Rather, much of the structure present in human language is indicative of apparent design *without* a designer. With that in mind, a different experimental methodology is offered – one which explains the emergence of linguistic structure as a result of cultural transmission, or iterated learning by multiple individuals.[2] The historical origins and theoretical viewpoints underpinning this approach are elaborated upon, and some recent experimental results obtained using this method are discussed. It shows that even in the absence of a communicative context, structural properties that are useful for communication can arise unintentionally. Finally, some directions for future research in this area are outlined.

2. Design without a designer

For centuries philosophers and linguists have debated the origins of linguistic structure and change in language. One of the central mysteries involves identifying the source of those changes and innovations that lead to increasing structure. The intuitive answer is of course us, the speakers of language. Yet whilst languages change and evolve as a result of differential patterns of usage among speakers, they do not do so as a result of any intentional design on an individuals' part. As Keller (1994) points out, we cannot analyse a historical change like the shift in word ordering from Object-Verb to Verb-Object in Middle English, and come to the conclusion that it is an instance of human design. He refers to events like this as 'phenomena of the third kind' – grouping together things that are neither man-made nor entirely natural, but which are instead "the result of human actions but not the goal of their intentions" (p. 56). He argues that we need to invoke an 'invisible-hand' explanation for language, adopting the metaphor proposed by the economist and philosopher Adam Smith to explain how locally self-serving actions of individual investors can unexpectedly lead to group-level prosperity. If this thesis is correct, it is only through developing an understanding of how apparent design emerges *without* a designer that we can hope to discover the origins of linguistic structure.

For Keller, who views language change as a special instance of sociocultural change, explaining this property of language means seeing it as a product of cultural evolution. It is obvious that language and culture are linked, but what does it mean to claim that language is a product of cultural evolution? Although ideas concerning cultural evolution have been around for as long as biological evolution, significantly less progress has been made in the former than the latter (Mesoudi et al., 2006). Whilst there is a good understanding of the origins and operations of many of the cognitive mechanisms underlying language that came about as a result of biological changes affecting our phylogeny, less is understood about the dynamics arising from language being culturally transmitted between individuals.

This is not a problem with our understanding of language in particular; but appears endemic to any culturally transmitted behaviour. Furthermore, it is not the case that every instance of cultural evolution requires an invisible hand explanation. If we look outwith human communication, we find many examples of culturally transmitted behaviours, such as tool-making and the kinds of incremental innovations we find in technological developments (Basalla, 1988; Petroski, 1992; Ziman, 2000), do seem to be directed and guided by human intentions, albeit that oftentimes the 'inventors' themselves cannot anticipate the eventual usage of the object to which they contribute some design feature. For some commentators (e.g. Hallpike, 1986; Pinker, 1997; Benton, 2000; Bryant, 2004), this intentional aspect is precisely what causes analogies between natural selection and cultural evolution to breakdown completely (Mesoudi, 2008). Yet instead of perceiving this as an either-or debate (cultural evolution either proceeds via intelligent human design or some 'blind' evolutionary process) Dennett & McKay (2006) encourage us to think of cultural change as "a continuum from intelligent, *mindful* evolution through to oblivious, *mindless* evolution." (italics original). They go on to claim that:

> "in cultural evolution...there are undeniable cases of cultural features that evolve by Darwinian processes without any need to invoke authors, designers, or other intelligent creators. Most obviously, languages – words and pronunciation and grammatical features – evolve without any *need* for grammarians, deliberate coiners, or other foresighted guardians of these cultural items." (p. 353).

So this brings us back to our central question – if some aspects of linguistic structure are led by this 'invisible-hand', is it possible to capture this phenomenon and investigate it in the laboratory? It could be argued that, in a sense, we have already seen the invisible-hand at work in some of the interactive construction studies discussed in the introduction.[3] This is complicated however, by the fact that there are many other processes at work which could arguably play a more significant role in the eventual emergence of structure.[4] Isolating exactly which elements arose through intentional design, and which through these more subtle and hidden forces may prove to be impossible in *any* experiment involving human participants. However, if we wish to learn the extent to which useful structure can arise in the absence of any intention to create it, we could do worse than to start with a theory of how such a blind mechanistic process might occur in the first place.

3. Iterated language learning

Almost everyone will agree that cultural evolution involves individuals within a group engaging in some kind of social learning with one another. One of the simplest and most general models of this kind of process is known as **iterated learning** (Kirby & Hurford, 2002). Put simply, iterated learning refers to the process whereby someone learns a behaviour by observing someone else performing that behaviour. Crucially, the person being observed must also have acquired that behaviour in the same way. This process is most commonly conceived of as a linear (vertical) transmission chain, with the output from each person's learning becoming the input for the next 'generation', although other population structures involving horizontal transmission are possible (see Mesoudi & Whiten (2008) for a review of experiments exploring different types of transmission chains in a non-linguistic setting).

Over the past decade or so, this iterated learning model has provided a framework for understanding cultural evolution in general, and language evolution in particular. The majority of this work has been undertaken using computational and mathematical models (e.g. see both Brighton et al. (2005), Kirby et al. (2008b), and references within) to explore what effect cultural transmission has on the structure of language. In spite of the many variations in the different models, two robust findings appear to hold whenever iterated language learning is at work. Firstly, languages become easier to learn over time, and secondly, they do so by becoming more structured. In other words, languages are adapting. To understand why this is the case, we need to look at iterated language learning in more detail.

Most simulations begin with a small population of agents with no initial language. For instance, in Kirby (2001), there are only two agents in the model at any one time – an adult 'speaker' and a child 'learner'. Learners acquire a set of mappings between strings of characters (signals) and pairs of features taking different values (meanings) by observing the signal-meaning pairs produced by adults. Initially, as there is no language in place, the first adult generates random, unstructured (or holistic) signals when prompted with a meaning. These are heard by the learner, who uses this data to induce its own representation of the system, before becoming the adult. At this point, a new learner appears, and the process repeats.

Learners induce their representations by storing the signal-meanings in a list, and then searching for possible generalisations over that data. Crucially, learners are only being trained on *a sub-set* of the total number of meanings in the language. We can think of this as a kind of bottleneck on transmission, one which mimics an aspect of the so-called 'poverty of the stimulus' that we know applies to real language acquisition (Smith, 2003). Namely, how we acquire an infinite language system on the basis of exposure to just a limited sub-set of the data. With this in mind, we will term this kind of transmission constraint a **data bottleneck**.

It transpires that this data bottleneck is of vital importance in explaining what happens to the languages over time. If this bottleneck is very wide (i.e. if learners are exposed to all, or nearly all, of the total meanings in the language) the languages do not change from their original random forms. On the other hand, if this bottleneck is very narrow (i.e. learners are exposed to just a few meanings) the languages become highly unstable. In the first case, signal-meaning pairs are just being memorised and passed along. In the second case, as only a few meaning-signal pairs are being transmitted between adjacent generations, each learner is forced to reinvent huge swathes of the system anew each time and the language cannot stabilise. Neither of these situations resemble a good model of linguistic transmission as we know it. However, if the bottleneck is neither too narrow nor too wide, something interesting starts to happen. The languages that were initially unstructured become compositional – the signals get decomposed into smaller units representing different aspects of the meanings, then recombined in some principled way to signify the meaning as a whole.

The emergence of compositionality is an adaptive response to the pressure of being transmitted through the data bottleneck. If meanings are encoded compositionally, they are far more likely to 'survive' transmission and be acquired by the next generation than if they are encoded holistically. This is because they can be reconstructed on the basis of fewer examples – they are more generalisable. It is not essential to see each and every meaning-signal pair to know what the signals are, we can reliably infer them based on the structure in the pairs we do see.

The presence of the data bottleneck triggers the **principle of linguistic adaptation**. This principle applies whenever language learners encounter imperfect information about the system they are trying to acquire. In such instances, cultural transmission becomes an adaptive process, causing languages to emerge that are seemingly optimised to the problem of being transmitted from person to person (Kirby et al., 2008b, pp. 89). It goes without saying that the adaptive solutions seen in the models do not come from intentional acts by agents. Put simply, agents are not equipped with reasoning or planning abilities, nor do they have the teleological goal of creating an 'optimal' language programmed into them. Many do not even have language-specific learning mechanisms. Yet in spite of this, spontaneous order emerges. Iterated language learning therefore seems a likely starting point for exploring the unintentional emergence of linguistic structure.

4. General methodology

Inspired by the early work of Esper (1925), who explored analogical change by using miniature languages, and more recently the work of Griffiths et al. (2008), who successfully turned a model of iterated learning in a non-linguistic domain into an experiment, work was undertaken to establish whether or not humans act in the same way as the simulated agents, by constructing an experimental version of iterated language learning (Kirby et al., 2008a). The results of this study are reviewed later, but the remainder of this section explores the framework used in more detail.

The general method involves each participant learning a small artificial ('alien') language composed of a finite set of meanings (pictures) that are paired with signals (strings of letters, or possibly sounds). Once a participant has acquired this language, they are tested and their answers used to provide the training input to the next participant, who forms another 'generation' in the chain. This process repeats until the desired number of generations is reached. Throughout, participants are asked only to reproduce the language as accurately as they can; the source of their training data is not revealed, and they have no way of knowing the experiment is investigating emergence.

There are three distinct phases involved: training, testing, and transmission. During the **training phase**, participants are shown a picture from the set, alongside the signal string it is paired with, and informed that this is the way in which the alien would describe that image in its own language. The task is to learn the descriptions associated with each image to the best of their abilities. Training occurs via a computer program, which randomises the order in which each signal-meaning pair is presented, ensures all training items are seen, and controls the length of time each training item is shown for. The key variables to consider here are the amount of training each participant receives (i.e. the number of passes over the data they are given), whether this training occurs in one continuous session or in blocks, and whether training blocks are structured in some way or randomised.

Once training is complete, we move onto the **testing phase**, where participants are shown each picture in turn and instructed to supply the missing description. The final test can be preceded by a series of practice tests in between training blocks, which introduces the possibility of feedback being provided to facilitate learning. This option is left unexplored for now, but is a potential avenue for future work. The final responses from the testing phase are then used to generate a new set of training stimuli for the next generation during the **transmission phase**. It is during this final stage, which happens 'offline' after the participant has left, that some of the most interesting parameters can be explored, including the transmission bottleneck. One of the advantages of the iterated language learning methodology is that it allows us to test very specific hypotheses about what occurs during language transmission by giving us complete control over what gets passed on. It is this aspect that affords iterated language learning more simulation-like qualities than is typical in non-iterated artificial language experiments.

For instance, if we wished to test the hypothesis that a preference for shorter strings led to compositional structure, during the transmission phase we could artificially select only those strings that met some (possibly dynamic) string-length threshold and ensure that only these items were propagated to the next generation.[5] By examining the resulting languages that arise from this process of artificial selection we can determine whether this hypothesis is valid. In this case we are running the procedure like a simulation. We build in a condition to see what the future outcome is, and can then refine our intuitions as a result. Alternatively, if we wish to test the hypothesis that human learners actually *have* a bias towards producing shorter strings, we can just run the experiment without any such manipulations and examine the average length of strings at the end of the chain. In this case, we are using the methodology to experimentally test whether such a bias currently exists or not. Both strategies can be useful depending on the questions one wants to answer.

These are not the only considerations that need to be kept in mind. One obvious factor we have yet to mention is how we begin this process. It is clear that the first participant needs a language to learn. There are several manipulations we can make here, which are again dependent on the kinds of questions we are interested in. For instance, if we wish to know whether a particular structural system can be stably transmitted, then we should give that system to the first participant and monitor whether it changes as a result of iterated learning. If however, we are interested in learning something about how linguistic structure emerges, we cannot initialise the chains with a fully structured system. Instead, we can use randomly generated signals. A simple method for constructing these is by concatenating CV syllables (drawn from a large but finite set) to form longer strings. This produces a set of signal strings, which whilst containing some regularities (owing to the fact that they are constructed from a finite syllable set) is still highly unstructured with regards to the meanings.

Further consideration must be paid to the design of the meaning-space – or rather, the stimuli we use to depict them. Meaning-spaces themselves can be structured or unstructured, reflecting regularities and co-occurrences in the real world, or a controlled and simplified world of our choosing. In all of the studies discussed later, the pictures come from a small and highly structured meaning space consisting of three different dimensions (motion,[6] colour and shape), each of which contains three different variables (e.g. bouncing, straight and spiraling; black, blue and red; circle, square and triangle). This $3 \times 3 \times 3$ design yields a total of 27 different possible combinations.

Attention also needs to be given to how we analyse the data from the study. In simulations, modelers have free access to the grammars formed by the agents over the course of the run, making descriptions and comparisons of the systems at different stages of their evolution relatively straightforward. Obviously this is not possible with human participants, and so alternatives must be found. At least two different types of measurement are required. Firstly, a method of analysing the similarity of languages *across* generations, and secondly a method of analysing the structural properties of languages *within* generations.

In order to calculate whether or not the languages are becoming easier to learn over time, there needs to be some measure of transmission error that compares adjacent generations and shows how much deviation there is between the two. If error is low, we infer the languages are being easily acquired. One way to do this is to calculate the mean edit distances of corresponding strings in each generation – that is, the number of substitutions, replacements and deletions required to turn string *a* into string *a'* (Levenshtein, 1966). For instance, if we wanted to assess the similarity between the strings '**wogi**' and '**wong**' we could calculate the amount of effort it would take to turn one into the other. In this case, we would need one insertion (**n**) and one deletion (**i**), giving an edit distance of 2. This figure can be normalised for string length, and then calculated for a whole language, producing a single number between 0 and 1 reflecting the degree of change between it and its predecessor.

The *within* generations measure is more complicated and needs to quantify the amount of structure within the language at each point in time. One way to do this is to use the *pairwise distance correlation,*[7] which calculates the extent to which similar signal-strings are used to express similar meanings. Just as we use edit distance to measure differences between signals, the same technique can be used for meanings. So whilst a 'red bouncing square' and a 'blue bouncing square' have an edit distance of 1, a 'black spiraling circle' and a 'blue horizontal triangle' have a distance of 3. The idea is that if the mappings between signals and meanings is structure-preserving (as would be the case, for instance, if the language was compositional), we should see a large positive correlation between these two sets of distances. If the mappings between the two are largely idiosyncratic (as would be the case for the initial unstructured languages), we would expect no correlation between the two distances. In order to establish whether the correlation is significant, and to compare our results across different languages, we must compute the z-score for the veridical correlation, and compare it to a large Monte Carlo sample of the same string data with a randomised alignment of meanings. This allows us to check whether we could have observed the mapping as a result of a random assignment of our meanings with those particular signals or not. A full description of both PDC and the transmission error outlined above can be found in Kirby et al. (2008a).

5. Recent studies

This section explores some recent studies illustrating the way in which the experiments work, and how some of the parameters interact with one another in interesting ways. These studies have been selected to demonstrate the kinds of empirical questions the methodology can address, with each one focusing in greater detail on the notion of the transmission bottleneck. There have been few attempts to rigorously define this concept in the literature, despite the fact that it is pressure to adapt to the constraints imposed on transmission that is the source of emergent structure in the model. The fact that it is 'constraints' plural should also not be forgotten. There are many potential bottlenecks on language transmission, often working simultaneously and not always in the same direction (Kirby, 2001; Hurford, 2002). This raises some interesting questions: are all constraints on transmission alike, and can we usefully study them in the laboratory?

The first experiment implements a data bottleneck of the kind mentioned previously, and attempts to replicate the computational findings concerning learnability and structure. This is contrasted with a second experiment showing similar results can be obtained by relying solely on the natural memory constraints of our learners. The third experiment explores what happens when we apply artificial selection for languages capable of expressing a larger proportion of the meaning-space. The fourth and final experiment extends the third by attempting to increase the early transmission fidelity by doubling the training. It will be seen that in all four cases the languages adapt over time to become more learnable and structured, often in interesting and unexpected ways.

5.1 The data bottleneck

The first experiment can be found in full in Kirby et al., (2008a). They attempted to replicate the computational finding that iterated learning leads to adaptation. The initial languages were randomly generated and used the same structured $3 \times 3 \times 3$ meaning-space described earlier. In addition, a bottleneck was placed on the amount of data each participant was trained on. Instead of having full access to all 27 meaning-signal during training, each participant was trained on exactly

14 of them, selected at random during the transmission phase at each generation. There were three blocks of training, during which each item was seen twice, for six seconds. 40 participants were recruited and offered £5 to take part. They were randomly assigned to one of four different language chains, each of which ran for ten generations.

The transmission error and structure scores are reproduced here in Figure 1. From these it is possible to confirm both predictions: the languages are adapting to become significantly more learnable and structured over time (as shown by a mean decrease in transmission error between first and final generations of 0.638, SD = 0.147; t(3) = 8.656; $P < 0.002$, and a mean increase in structure of 5.578, SD = 2.968, t(3) = 3.7575, $P < 0.02$).

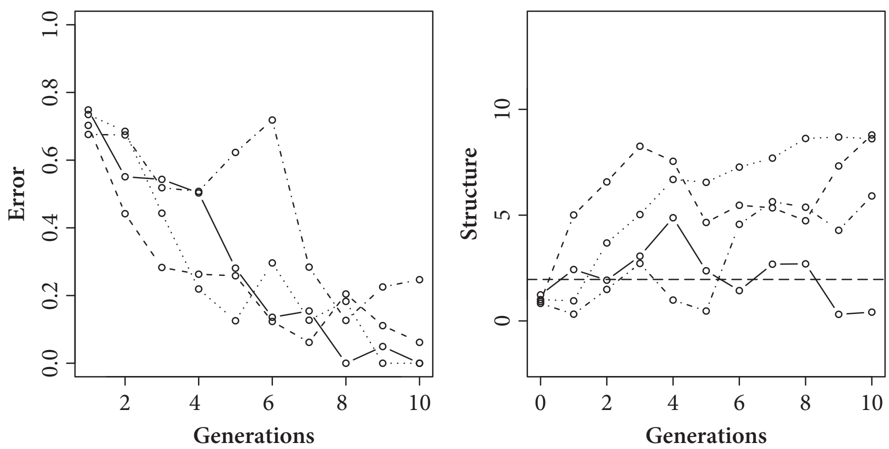


Figure 1. Graphs showing the normalised error (left) and pairwise distance correlation (right) scores by generation, of four transmission chains when a data bottleneck was present. These results show that the systems are all evolving to become more learnable and more structured over time. Points above the dotted line in Fig 1 (right) represent significant structural regularities between signal-meaning mappings. Re-drawn from Kirby et al. (2008a) with permission

When analysing the languages qualitatively, Kirby et al. (2008a) made an interesting discovery. Whilst the languages that arose in this condition in the computational models were all compositional in structure, the languages in the experimental version were not. Instead, every chain showed a massive decrease in the number of distinct strings being used, introducing widespread underspecification. Reduction in the number of words to remember could not account for the low error values alone. To get any figure below 50% there must be inter-generational agreement on those items that were *not* seen during training, and this could only happen if the signals were structured. Table 1 shows the language from the chain with the highest score[8] at generation 10. From this we can see that although certain meanings are now underspecified, there is still a systematic relationship between meanings and signals. For instance, all horizontally moving shapes are called 'tuge', regardless of shape or colour.

Table 1. Table showing the language with the highest structure score at generation 10 in the data bottleneck condition. Signals are located in cells corresponding to their meaning features. Columns align with colours, whilst motion and shape features inhabit rows. This particular language exhibits systematic underspecification, which is a successful strategy enabling learners to reproduce the whole language from just a fragment. Re-drawn from Kirby et al. (2008a) with permission

	black	**blue**	**red**	
	miniku	miniku	miniku	circle
bounce	tupim	tupim	tupim	square
	tupin	tupin	tupin	triangle
	tuge	tuge	tuge	circle
horizontal	tuge	tuge	tuge	square
	tuge	tuge	tuge	triangle
	poi	poi	poi	circle
spiral	poi	poi	poi	square
	poi	poi	poi	triangle

Clearly this is a somewhat unexpected result. Nevertheless, it tells us two important things: (1) the dynamics of cultural transmission certainly give rise to adaptive structure in a laboratory setting, and (2) something more than a pressure for easy transmission must be required to explain the emergence of compositionality. This second idea will be explored later, but before that we should discuss the data bottleneck in more detail. The data bottleneck works by forcing perceived structural patterns in previously encountered stimuli to be generalised to novel stimuli. Even if these perceived patterns occurred only through chance sampling of the data, the data bottleneck works to massively amplify their effect throughout the whole language (Cornish et al., *submitted*). What was previously just an inference of structure before learning, is now a *bona fide* instance of it after learning. This is particularly the case in earlier generations when there is more variation in the system.

The next experiment explores what happens when we remove the data bottleneck completely.

5.2 The memory bottleneck

The idea of removing the data bottleneck may seem strange given the computational findings. For instance, Smith (2003) clearly shows that running an iterated learning model with no data bottleneck results in no cultural evolution taking place at all. Yet there is an important difference between idealised computational agents and human participants: the agents in many of these models are perfect learners. They memorise meaning-signal pairs flawlessly. Given the fact that human memory is not this reliable, it might be worth investigating whether this 'memory bottleneck' could play a role on the emergence of linguistic structure.

This memory bottleneck was also at work during Kirby et al.'s (2008a) study. A problem arises in that there is a potential confound between the two types of bottleneck. According to the principle of linguistic adaptation discussed earlier, cultural transmission only becomes adaptive when the learner is presented with imperfect information. In order to assess what, if anything, imperfect learning has contributed to the overall result in the earlier study, it is necessary to first remove this confound by eradicating the effect of the imperfect data. With that in mind the experiment was re-run, this time with participants given full access to all 27 meanings during training. The training and testing phases were held proportional to the previous experiment; although there were more training items, each was seen the same number of times and for the same duration as before.

The results of this experiment are shown in Figure 2. As before, the transmission error significantly decreases over the course of the experiment indicating that the language is becoming easier to learn (mean decrease of 0.446, $SD = 0.193$, $t(3) = 4.628$, $p < 0.009$). Interestingly, comparing the initial error values with those in the data bottleneck condition in Figure 1 we find that there is no significant difference between the two. This suggests that removing the data bottleneck has not made the task any easier or harder for the participants. Looking at the structure scores, we also find that they significantly increase over time as well (mean increase of 7.396, $SD = 1.629$, $t(3) = 9.079$, $p < 0.001$).

In fact, if we compare Figures 1 and 2 there do not appear to be any real differences in development patterns of the languages at all.[9] This is in fact surprising, as we might expect that once some structure had emerged in the systems, having full access to the data may actually facilitate acquisition and lead to more overall stability. This is inferred on the basis that it is easier to detect structural regularities when provided with more evidence than it is with less. The fact that we only have four chains means that we cannot rule out the possibility that we have just been unlucky not to observe more stability in this instance. However, it could also be the case that there are features within the language itself which prevents such stability emerging.

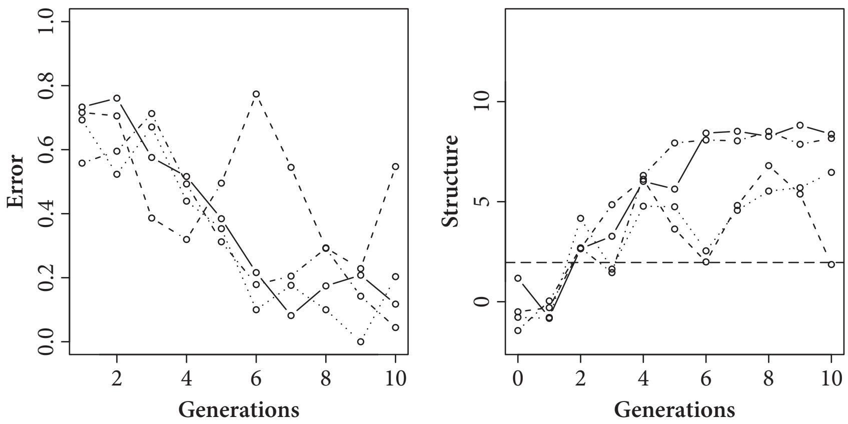


Figure 2. Graphs showing the normalised error (left) and pairwise distance correlation (right) scores by generation, of four transmission chains when no data bottleneck was present. These results show that imperfect learning of data achieves the same result as exposure to novel stimuli. The systems are all evolving to become more learnable and more structured over time. Points above the dotted line in Fig 2 (right) represent significant structural regularities between signal-meaning mappings

We can analyse the languages qualitatively to get a better impression of what is going on. Table 2 shows the language from the chain with the highest structure score at generation 9. We can see that this language also exhibits systematic underspecification. Unlike the previous example however, there appears to be tentative evidence for some internal structure – that is, some of the signals begin to look like they are decomposable into sub-units. For instance, signals used for spiraling objects can be broken up into a prefix (wag-, nuak- or wagin-) which represents the shape, and a suffix (-ini) which we could gloss as meaning 'spiraling'. However, these prefixes only seem to apply locally. They are not used elsewhere.

There also seem to be several irregulars present – for instance, 'mucapo' and 'nukapo'. Examining the generations immediately before and after, it is apparent that these forms have persisted a while, but are not stably associated with specfic meanings. To borrow an analogy from phonology, they seem to be in free variation. It seems that these alternating variants do disappear eventually, but only gradually. Their presence could explain why we do not see as much stability here as in the data bottleneck condition however: having full access to the data allows irregulars to survive, but complicates the acquisition process by making it necessary to memorise these exceptions on a case by case bases.

Table 2. Table showing the language with the highest structure score at generation 9 in the no data bottleneck condition. Signals are located in cells corresponding to their meaning features. Columns align with colours, whilst motion and shape features inhabit rows. This language also exhibits systematic underspecification, although there appear to be signs of internal structure indicating shape and motion amongst spiraling objects, and irregulars (e.g. 'mucapo' and 'nukapo')

	black	**blue**	**red**	
	nucapo	nucapo	nucapo	circle
bounce	nucapo	mucapo	mucapo	square
	nucapo	nukapo	mucapo	triangle
	hapo	hapo	hapo	circle
horizontal	hapo	hapo	hapo	square
	hapo	hapo	hapo	triangle
	wagini	wagini	wagini	circle
spiral	nuakini	nuakini	nuakini	square
	waginini	waginini	waginini	triangle

The results of these two experiments show us two important things. Firstly, adaptive systems that could be used for communication can emerge in an experimental setting, without a designer. It is worth stressing this point again – the participants involved are not 'solving' these transmission 'problems'. In fact, they are not even aware that there *are* transmission problems. Secondly, the results of our second experiment extend previous work by showing that the presence of a data bottleneck is not essential for cultural transmission to become adaptive. Instead, the key is imperfect information. The source of that imperfect information (lack of exposure to data, human memory limitations) appears irrelevant.

Next we move onto exploring a topic that was hinted at earlier – the emergence of compositionality.

5.3 Another kind of bottleneck: Forcing expressivity

One of the reasons why underspecification is so prevalent in the previous experiments is that the presence of homonyms creates a snowball effect: once one appears in the system it sends a strong signal to later learners, encouraging more to emerge. Given the fact that the only bottlenecks we have explored so far demand nothing more than the system be learnable, it could well be the case that this represents the ideal solution for the language; after all, the most learnable system is one in which there is just one name for everything.[10] However, what would happen if the task were to change slightly, requiring that the language not only be learnable, but also expressive (i.e. be capable of uniquely expressing more of the meanings)?

In the simulation literature, expressivity is enforced as a matter of course. Although never explicitly addressed or discussed as a type of bottleneck, every model has an expressivity requirement built into it somewhere. In most cases it is implicit in the learning or production mechanisms of the agents, built in to model the known one-to-one mapping bias humans possess (Smith, 2003). Fortunately however, this means there are well-understood techniques for enforcing expressivity that can be borrowed from the models. The simplest of these involves filtering out repeated instances of the same signal being attached to multiple meanings. The first novel signal-meaning pair produced by generation *n-1* is propagated into the training input to generation *n*. Thereafter any repeats of those signals are removed. This ensures that the input to generation *n* only ever contains one-to-one mappings between signals and meanings.

Kirby et al. (2008a) ran an iterated language learning experiment using just such a filtering method. As before they used four randomly generated languages to initialise four distinct chains. Once the results of the final test were collected, 14 new items were sampled at random for the next generation to train upon and homonyms were removed. The error and structure scores are reproduced in Figure 3. These show that once again, the language is adapting to become significantly more learnable (shown by a mean decrease in transmission error between first and final generations of 0.427, SD = 0.106; t(3) = 8.0557; $P < 0.002$), and more highly structured over time (shown by a mean increase in structure of 6.805, SD = 5.390, t(3) = 2.525, $P < 0.05$).

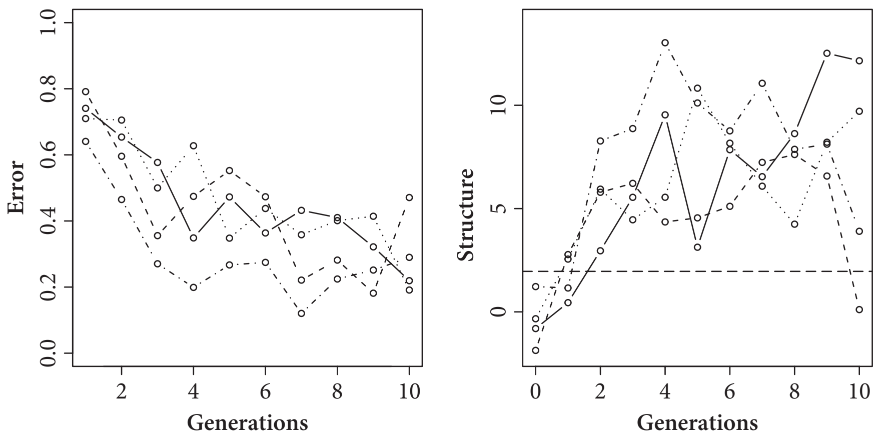


Figure 3. Graphs showing the normalised error (left) and pairwise distance correlation (right) scores by generation, of four transmission chains when both a data bottleneck and a homonym filter was present. These results show that despite the blocking of underspecification, structure is still emerging leading to the language to become increasingly learnable. Points above the dotted line in Fig 3 (right) represent significant structural regularities between signal-meaning mappings. Re-drawn from Kirby et al. (2008a) with permission

Table 3 shows the language with the highest structure score at generation 9. From this it appears we have found a compositional system at last. Although not perfectly regular, each signal seems to be composed of three morphemes, each corresponding to a different feature of meaning. For instance, the first letter corresponds to colour, and there are regular suffixes indicating motion. Shape is the most irregular, and seems to be encoded by remaining letters. It should be noted that the occurrence of this degree of compositionality is both rare and fleeting – it features in just two of the four chains, and does not appear to be stable.[11] Nevertheless, the fact that it emerges at all is encouraging, and also serves to drive home the point once and for all that the appearance of structure in these studies is an invisible-hand process. As Kirby et al. (2008a) are at pains to point out, as far as participants are concerned, the filtering bottleneck is an invisible modification. The individuals involved would have had no way of knowing which condition they were in, and yet the kind of languages they produced differed radically.

Table 3. Table showing the language with the highest structure score at generation 9 in the filtering condition. Signals are located in cells corresponding to their meaning features. Columns align with colours, whilst motion and shape features inhabit rows.This particular language exhibits signs of compositionality, with signals being composed of three morphemes representing colour, shape and motion respectively. Of these, only colour and motion are consistent. Note also the presence of an irregular, 'renana'. Re-drawn from Kirby et al. (2008a) with permission

	black	**blue**	**red**	
	nehoplo	lahoplo	rehoplo	circle
bounce	nereplo	laneplo	replo	square
	nekiplo	lakiplo	rahoplo	triangle
	neheki	lahoki	reneki	circle
horizontal	nereki	lereki	renana	square
	nekeki	lakeki	raheki	triangle
	nehopilu	lahopilu	repilu	circle
spiral	nepilu	lanepilu	repilu	square
	nekipilu	lakipilu	rahopilu	triangle

Perhaps one of the reasons why compositionality does not stabilise after it emerges is due to the comparatively extreme learning conditions imposed upon participants. We can see by looking at the transmission error levels obtained by the first generation in all three studies so far that participants are struggling to accurately learn the items they are trained on. In most cases they are only learning between 20–35 per cent of all meaning-signal pairs, translating to a 40–70 per cent accuracy on seen items. This is hardly surprising given that they see each item only six times in total. This begs the question of how important early transmission fidelity is. Would increasing the amount of exposure to each training item lead to a more stable compositional language? This is the question posed in our final study.

5.4 Increasing early transmission fidelity

This study followed the same outline of the filtering condition, but with one slight alteration; each training item appeared twice as often. The main aim of this double training condition was to see whether or not a compositional system could be stably transmitted once it emerged. Figure 4 shows the error and structure scores of the resulting chains. As we would expect, transmission error and structure are both significant (mean decrease in error of 0.35, SD = 0.063; $t(3) = 11.079$, $p < 0.0008$ and mean increase in structure of 9.83, SD = 2.639; $t(3) = 7.449$, $p < 0.003$). Again, as we would expect, comparing Figures 3 and 4 the transmission error scores of the first generation are much lower in the double training condition. Furthermore, it appears at least one of the chains results in a language that remains relatively stable for at least five generations.

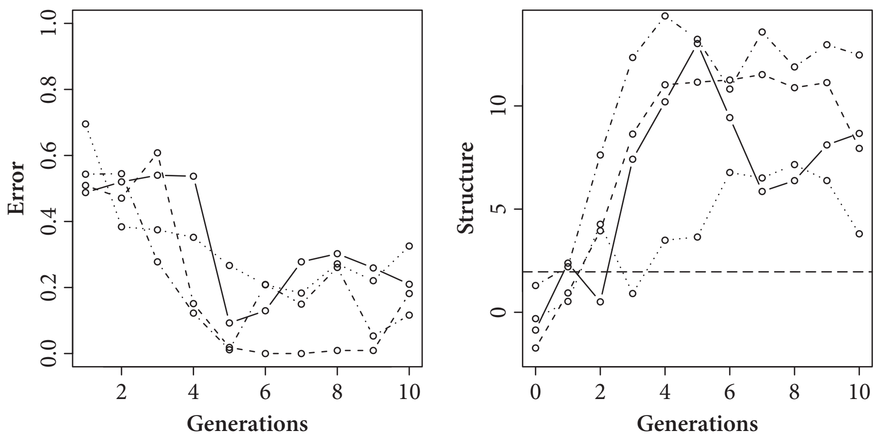


Figure 4. Graphs showing the normalised error (left) and pairwise distance correlation (right) scores by generation, of four transmission chains in the double training condition. These results show that once more the systems are adapting to become more learnable and more structured over time. Periods of high stability were also observed. Points above the dotted line in Fig 4 (right) represent significant structural regularities between signal-meaning mappings

Table 4 shows the structure of this stable chain, taken at the mid-point in generation 7. Interestingly enough, this system seems to combine clear elements of a compositional *and* an underspecified system. Each signal is composed of two parts – a prefix describing the colour, and a suffix describing the motion (black objects are referenced by a null morpheme). How is this possible, when the languages are still being filtered for homonyms? The answer is deceptive. As long as homonyms are evenly distributed throughout the language, and as long as there is a very small amount of compositionality allowing the chance to reconstruct a form if it is unlucky enough to not be selected for transmission, the filtering process can be easily bypassed. In this case, the language has perfectly adapted to find this delicate equilibrium, resulting in an unexpected but highly elegant solution.

Table 4. Table showing a stable language, taken at generation 7 in the double training condition. Signals are located in cells corresponding to their meaning features. Columns align with colours, whilst motion and shape features inhabit rows. This particular language exhibits properties of both systematically underspecified systems and compositional ones. Colours and motions are signified using distinct morphemes – although black objects are signified with a null morpheme – whilst the shape dimension remains underspecified. This system appears in spite of filtering to remove homonyms

	black	**blue**	**red**	
	gahili	pagahili	megahili	circle
bounce	gahili	pagahili	megahili	square
	gahili	pagahili	megahili	triangle
	linu	palinu	melinu	circle
horizontal	linu	palinu	melinu	square
	linu	palinu	melinu	triangle
	wenu	pawenu	mewenu	circle
spiral	wenu	pawenu	mewenu	square
	wenu	pawenu	mewenu	triangle

As if that were not impressive enough, the same solution was found again, only this time it emerged in a chain that was previously fully compositional. The two instances are shown in Tables 5 and 6, and refer to the chain with the highest structure in generation 4. As ever we cannot make any strong claims about the emergence of a particular structure given that we only have four data points as a baseline, but this is suggestive of the fact that fully compositional systems are perhaps still too difficult to maintain even with extra training. Further study is required here to fully assess the difficulty of the task.

Table 5. Table showing the language with the highest structure at generation 4 in the double training condition. Signals are located in cells corresponding to their meaning features. Columns align with colours, whilst motion and shape features inhabit rows. This language is fully compositional, with each meaning feature being consistently encoded using a distinct morpheme

	black	blue	red	
	wakeki	hunkeki	pokeki	circle
bounce	wakiki	hunkiki	pokiki	square
	wanikuko	hunikuko	ponikuko	triangle
	wakemo	hunkemo	pokemo	circle
horizontal	wakimo	hunkimo	pokimo	square
	waknimo	hunimo	ponimo	triangle
	wakekuko	hunkekuko	pokekuko	circle
spiral	wakikuko	hunkikuko	pokikuko	square
	wanikuki	hunikuki	ponikuki	triangle

Table 6. Table showing the same language as Table 5 at generation 10. Signals are located in cells corresponding to their meaning features. Columns align with colours, whilst motion and shape features inhabit rows. The language that was previously fully compsitional has become mixed – incorporating features of both systematic underspecification and compositionality. Colours and motions are signified using distinct morphemes whilst the shape dimension remains underspecified

	black	blue	red	
	wakiko	pokiko	pokiko	circle
bounce	wakiko	hekiko	pokiko	square
	wakiko	hekiko	pokiko	triangle
	wanimo	henimo	ponimo	circle
horizontal	wanimo	henimo	ponimo	square
	wanimo	hekiko	ponimo	triangle
	wahikeko	hehikeko	pohikeko	circle
spiral	wahikeko	hehikeko	pohikeko	square
	wahikeko	hehikeko	pohikeko	triangle

This section began by asking a question about whether all constraints on transmission were alike, and whether they could be usefully studied in the laboratory. From the data we have seen it appears that the answer to the first part is no. While it seems that the data and the memory bottlenecks can be functionally classified as the same kind of constraint – one which forces languages to be learnable – the filtering bottleneck described in the third experiment appears to play a different kind of role, by forcing the system to also be minimally expressive. The results of the final experiment are harder to interpret, but do suggest more work needs to be undertaken to explore the different kinds of adaptations that emerge in experiments like these. Certainly it acts as a reminder of the many surprising, and often unintuitive, ways in which the system can evolve during transmission. Given the fact that we are in a position to answer the first part of our question after performing these studies, logic suggests that the answer to the second part must be yes. However, we should not feel too content. There is work to be done.

6. Future directions

This paper has touched upon several different ideas. The first is that structural changes and innovations in natural language are typically not the result of intentional actions designed to bring about that goal. It is important to remember this if we want to develop a full picture of how linguistic structure emerges. The second idea is a suggestion to help handle the first. By taking a cultural evolutionary perspective on language, we can develop new methods that allow us to investigate the appearance of apparent design without a designer. One such method is the iterated language learning framework, which has been recently developed into an experimental methodology revolving around the repeated cultural transmission of simple 'alien' languages through the minds of participants.

Work in this area is still in its infancy. This paper outlined four recent studies that looked a little closer at the notion of transmission bottlenecks, but there remain many avenues still left unexplored. These include:

1. *the design of the meaning space*: the emerging linguistic structures are obviously highly dependent on the structure of the meaning spaces that they are evolving to express, and as such, having a more realistic model of the world is an obvious area for improvement. Some of the groundwork has already been covered mathematically in simulations (Kirby, 2007), and to some extent explored in robotic agents who evolve their own meaning spaces over time (Steels, 2003), but as yet this area still represents ongoing work in our research lab and others.
2. *different population structures*: possible population structures can be defined by three different parameters – the size of the population, the direction of transmission (vertical, horizontal or mixed) and the network structure dictating who learns from whom. In this paper we have only examined

the simplest population structure imaginable – a single vertically transmitted chain – although work expanding this is ongoing.

3. *iterated language learning in children:* one obvious extension is to explore this process in children, as they have long been implicated in the emergence of language. Work in this area would contrast nicely with the recent non-linguistic diffusion chain work being undertaken with children (Flynn, 2008).
4. *different modalities:* another area which could be interesting is to use spoken or gestural signals rather than written. Some work has recently been undertaken exploring the iterated learning of musical tones (Brown & Tamariz, *submitted*) in musicians vs. non-musicians, with scope for extending this work further.

The rise in the number of studies exploring the emergence of human communication experimentally is deeply encouraging. It is important that this work continues, but that in future it focuses on all edges of the cultural evolutionary continuum: on the emergence of systems that arise through intentional human design, on systems that arise unintentionally through vertical and horizontal transmission, and systems that arise through combinations of the two.

Notes

1. Also implicit in each of these experiments is the notion that much of the character of these systems arise from social interactions between individuals, and do not just emerge directly from the underlying cognitive systems of those who possess them – hence why there is a need to study this aspect of the process in the first place.

2. Strictly speaking, we are not witnessing the evolution of communication systems in these studies, but the evolution of signs. This is because there is no actual communication taking place between participants; the task is all about learning a system, and not about using it for anything. The distinction is important because, as we will see later, although we find structural features emerging that *are* useful for communication, we sometimes find these systems developing in ways we would not expect if they were being used communicatively.

3. The author would like to thank both Bruno Galantucci and Simon Garrod for making this point clear. While interactions between participants may involve some reasoning and purposeful design, the negotiation process is also a complex dynamic system at work. As such, it has invisible-hands of its very own; shaping, guiding and prompting structure into being. This idea would help explain why the creation of a successful system is never guaranteed in these studies.

4. This is not intended as a criticism of the interactive construction methodology. We should be mindful of Dennett & McKay's continuum of cultural evolution here, and the need to explore all lengths of it.

5. It should be remembered that studying processes of artificial selection (e.g Mendel's peas, the selective breeding programs employed by farmers, etc.) were what led to the breakthroughs in understanding how biological evolution worked. One of the points being arguing for here is that similar tactics of studying artificial selection in language and other culturally transmitted behaviours can lead to similar advances in understanding cultural evolution. This is consistent with the agenda laid out in Mesoudi et al. (2006).

6. Motion was represented using a directional arrow, although real movement could be achieved using video instead of static images.

7. This is the measure of structure used in Kirby et al. (2008a), although it is not named as such. The main advantage of the PDC method is that it is detects all kinds of structure – not just compositionality. However, there are other methods for assessing structure within generations that may be more suited for analysing the emergence of compositionality, such as RegMap (Tamariz & Smith, 2008; Cornish et al., 2009).

8. It should be noted that these examples are not always representative of the range of results found in each condition and should be treated in the spirit with which they are offered – as individual case-studies. The complete data for every study is available on request: hannah@ling.ed.ac.uk

9. It should be noted that the striking peaks in transmission error seen at generation 6 in different chains in Figs. 1 and 2 appear coincidental, and simply an unfortunate reminder of how fragile single transmission chains are to individual variations in recall ability.

10. Obviously this system would not be at all useful for communication.

11. In this regard at least, iterated language learning shares similarities with the communication game experiments described in the Introduction. Whilst compositionality is a ubiquitous feature of human language, it emerges rarely in experiments investigating the novel emergence of communication systems.

References

Basalla, G. (1988). *The evolution of technology.* Cambridge: Cambridge University Press.

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

Benton T (2000). Social causes and natural relations. In Rose, H., & Rose, S. (eds) *Alas, Poor Darwin*, pp. 206–224. New York: Harmony.

de Boer, B. (2001). *The Origins of Vowel Systems.* Oxford: Oxford University Press.

Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, 2(3): 177–226.

Brown, E., & Tamariz, M. (*submitted*) The effect of extensive training on the expectation and generalisation of regularity in artificial musical language learning.

Bryant, J.M. (2004). An evolutionary social science? A skeptic's brief, theoretical and substantive. *Philosophy of the Social Sciences*, 34:451–492.

Christiansen, M.H. (2000). Using artificial language learning to study language evolution: Exploring the emergence of word order universals. In J.L. Dessalles & L. Ghadakpour (Eds.), *The evolution of language: 3rd international conference* (pp. 45–48). Paris: Ecole Nationale Supérieure des Télécommunications.

Cornish, H., Tamariz, M., & Kirby, S. (2009). Complex adaptive systems and the origins of adaptive structure: what experiments can tell us. *Language Learning*, 59(S1): 187–205.

Dennett, D. & McKay, R. (2006). A continuum of mindfulness. *Behavioral and Brain Sciences*, 29(4), 353–354.

Esper, E.A. (1925). *A technique for the experimental investigation of associative interference in artificial linguistic material*. Philadelphia: Linguistic Society of America.

Flynn, E. (2008). Investigating children as cultural magnets: do young children transmit redundant information along diffusion chains? *Philosophical Transactions of the Royal Society B: Biological Sciences,* 363(1509): 3541–3551.

Galantucci, B. (2005). An Experimental Study of the Emergence of Human Communication Systems. *Cognitive Science*, 29(5): 737–767.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialog—A study in conceptual and semantic coordination. *Cognition, 27,* 181–218.

Garrod, S., & Doherty, G. (1994). Conversation, coordination and convention—An empirical-investigation of how groups establish linguistic conventions. *Cognition, 53,* 181–215.

Garrod, S., Fay, N., Lee, J., Oberlander, J, & Macleod, T. (2007). Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cognitive Science*, 31(6): 961–987.

Goldin-Meadow, S. & Mylander, C. (1983). Gestural communication in deaf children: Non-effect of parental input on language development. *Science*, 221(4608): 372–374.

Griffiths, T.L. & Kalish, M.L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, 31(3): 441–480.

Griffiths, T.L., Kalish, M.L., & Lewandowsky, S. (2008). Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 363(1509): 3503–3514.

Hallpike, C.R. (1986). *The principles of social evolution*. Oxford: Clarendon Press.

Healey, P.G.T., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. *Cognitive Science*, 31(2): 285–309.

Hudson-Kam, C.L. & Newport, E.L. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development,* 1(2): 151–195.

Hurford, J. (2002). Expression/induction models of language evolution: dimensions and issues. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.

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

Kirby, S., & Hurford, J. (2002). The Emergence of Linguistic Structure: An overview of the Iterated Learning Model. In Angelo Cangelosi and Domenico Parisi, editors, *Simulating the Evolution of Language*, pp. 121–148. London: Springer Verlag.

Kirby, S., Dowman, M., & Griffiths, T.L. (2007). Innateness and culture in the evolution of language. *PNAS*, 104(12): 5241–5245.

Kirby, S., Cornish, H., & Smith, K. (2008a). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS*, 105(31): 10681–10686.

Kirby, S., Smith, K., & Cornish, H. (2008b). Language, Learning and Cultural Evolution: How linguistic transmission leads to cumulative adaptation. In R. Cooper & R. Kempson, editors, *Language in Flux: Dialogue Co-ordination, Language Variation, Change and Evolution,* pp. 81–108. London: College Publications.

Keller, R. (1994). *On language change: The invisible hand in language.* London: Routledge.

Levenshtein, V.I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics-Doklady,* 10: 707–710.

Mesoudi, A. (2008). Foresight in cultural evolution. *Biology and Philosophy*, 23(2): 243–255.

Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B*, 363, 3489–3501.

Mesoudi, A., Whiten, A. & Laland, K.N. (2006). Towards a unified science of cultural evolution. *Behavioral and Brain Sciences,* 29(4), 329–383.

Oudeyer, P-Y. (2006). *Self-Organization in the Evolution of Speech.* Oxford: Oxford University Press.

Petroski, H. (1992). *The evolution of useful things.* New York: Vintage Books.

Pinker, S. (1997). *How the mind works.* New York: Norton.

Pickering, M.J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169–225.

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, T.C., Kirby, S., & Ritchie, G.R.S. (2009). Signalling signalhood and the emergence of communication. *Cognition, 113*(2), 226–233.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *PNAS*, 104(18): 7361–7366.

Senghas, A., & Coppola, M. (2001). Children creating language: how Nicaraguan sign language acquired a spatial grammar. *Psychological Science*, 12: 323–328.

Smith, K. (2003). Learning biases for the evolution of linguistic structure: an associative network model. In *ECAL03*, pp. 517–524.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Science.* 7(7): 308–312

Tamariz, M., & Smith, A.D.M (2008). Regularity in Mappings Between Signals and Meanings. In A.D.M. Smith & K. Smith and R. Ferrer-i-Cancho (Eds), *Proceedings of the 7th International Conference on the Evolution of Language,* pp. 315–322. World Scientific.

Vogt, P. (2005). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence,* 176: 206–242.

Wonnacott, E., & Newport, E.L. (2005). Novelty and regularization: The effect of novel instances on rule formation. In A. Brugos, M.R. Clark-Cotton, & S.Ha (eds.), *BUCLD 29: Proceedings of the 29th Annual Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press.

Ziman, J. (Ed.) (2000). *Technological innovation as an evolutionary process.* Cambridge: Cambridge University Press.

Biographical note

Hannah Cornish is an ESRC funded postgraduate student working at the Language Evolution and Computation Research Unit at the University of Edinburgh. Her Ph.D. supervisors are Prof. Simon Kirby and Dr. Monica Tamariz.