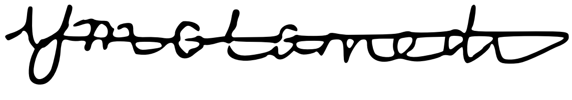
Artificial sign language learning: a method for evolutionary linguistics

Yasamin Motamedi-Mousavi

Doctor of Philosophy
Centre for Language Evolution
The University of Edinburgh
2017

Declaration of own work

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, and that this work has not been submitted for any other degree or qualification.



Yasamin Motamedi-Mousavi

Abstract

Previous research in evolutionary linguistics has made wide use of artificial language learning (ALL) paradigms, where learners are taught artificial languages in laboratory experiments and are subsequently tested in some way about the language they have learnt. The ALL framework has proved particularly useful in the study of the evolution of language, allowing the manipulation of specific linguistic phenomena that cannot be isolated for study in natural languages. Furthermore, this framework can test the output of individual participants, to uncover the cognitive biases of individual learners, but can also be implemented in a cultural evolutionary framework, investigating how participants acquire and change artificial languages in populations where they learn from and interact with each other.

In this thesis, I present a novel methodology for studying the evolution of language in experimental populations. In the artificial sign language learning (ASLL) methodology I develop throughout this thesis, participants learn manual signalling systems that are used to interact with other participants. The ASLL methodology combines features of previous ALL methods as well as silent gesture, where hearing participants must communicate using only gesture and no speech. However, ASLL provides several advantages over previous methods. Firstly, reliance on the manual modality reduces the interference of participants' native languages, exploiting a modality with linguistic potential that is not normally used linguistically by hearing language users. Secondly, research in the manual modality offers comparability with the only current evidence of language emergence and evolution in natural languages: emerging sign languages that have evolved over the last century. Although the silent gesture paradigm also makes use of the manual modality, it has thus far seen little implementation into a cultural evolutionary framework that allows closer modelling of natural languages that are subject to the processes of transmission to new learners and interaction between language users.

The implementation and development of ASLL in the present work provides an experimental window onto the cultural evolution of language in the manual modality. I detail a set of experiments that manipulate both linguistic features (investigating

category structure and verb constructions) and cultural context, to understand precisely how the processes of interaction and transmission shape language structure. The findings from these experiments offer a more precise understanding of the roles that different cultural mechanisms play in the evolution of language, and further builds a bridge between data collected from natural languages in the early stages of their evolution and the more constrained environments of experimental linguistic research.

Lay summary

Humans show a unique ability in the animal kingdom to learn complex systems that allow us to communicate with an unlimited expressive power. Languages are structured in such a way that allows this expressive power, whilst limiting the demands it places on our memory and processing abilities; they re-use and re-combine elements in language to signal similarities and dissimilarities across meanings. For example, the re-use of *blue* in the phrases *blue shoes* and *blue chair* signals a shared attribute across these meanings – that they are both blue. Such structure is prevalent across human languages, and enables us to talk about an uncountable range of objects, states and events.

This thesis is concerned with understanding how this structure emerges, through language learning and use. How does the use of language in communication affect the structure of language? And how does our need to learn language affect linguistic structure? I test these questions using experiments in which participants innovate gestural communication systems and use these systems in communication and learning scenarios. The use of gesture has two principal benefits: a) it reduces the influence of participants' existing linguistic knowledge, and b) it offers comparability with the only evidence we currently have of newly created languages (emerging sign languages that have evolved over the last 100 years). This method, of artificial sign language learning, allows the close and precise study of how language use in communication and language learning affects the evolution of linguistic structure.

The results that have been evidenced in this experimental work highlight the importance of language use and learning in the evolution of linguistic structure. Both processes provide pressures that shape language, communication leading to efficient and mutually comprehensible systems, and learning leading to simpler systems that are facilitate the learning process. These findings offer a unique insight into the precise roles that learning and use play in the evolution of language, and also provide data that complements and supports data collected from languages used in the real world.

Acknowledgements

For those who know me well, this uncharacteristic outpouring of emotion may be too much for you to handle. I suggest you look away now.

The Carnegie Trust was kind enough to fund the research that makes up this thesis, and have kept me fed and housed for the last three years, for which I am very grateful.

All three of my supervisors have gone far above and beyond their titular duties, and it is only a shame that more students do not have supervisors like them. As a team they have encouraged me every step of the way, and given me the notion that I occasionally do intelligent work. Simon's door has always been open to me, for the big problems and the characteristically Yasamin anxieties (of which there were many). He has put up with more eye-rolls and severe looks than any person in a position of authority sensibly would, or should, and frequently laughs at the jokes I make at his expense. Kenny has also often been assaulted, by puns rather than insults, but does not appear to hold it against me. If anything, he has gone out of his way to encourage me to develop as a student, and is in fact to blame for encouraging me to apply to do a PhD in the first place. Marieke has brought more than her fair share of careful consideration to my supervision, and her incredible ability for measured insight has frequently transformed how I have thought a piece of work. More importantly, her kindness and support has championed me through the most stressful parts of this experience. I must also add a special thanks to Jenny Culbertson, who, though she has not been an official supervisor, has been ever-present throughout this process, and whose critical (but kind) insight has been instrumental in refining my research.

The CLE as a whole has given me the most amazing PhD experience, giving me free reign to be my bizarre, grumpy self. That they allowed me to accost a bunch of unsuspecting new students with a D:Ream-soundtracked powerpoint presentation (complete with firework animations) says a lot about how accepting an environment the CLE provides. The people of the 1.15 office have also made up my harmonious academic family, and a few people in particular have really made the last few years for me: my desk-mate and head cheerleader, Carmen, the ever-compassionate Small Marieke, my biscuitty soulmate E, Steph (and Josephine), and Jon, Vanessa, Catriona, and Ashley, to name just a few. Repetition of the mantra ”if Kevin can do it, so can I” helped this thesis to its final state, so I must also thank my Number One Enemy, Kevin Stadler, for that useful motivation. I also have to thank Jasmeen for her yoga classes, which kept me sane and stretchy, and should really be an essential part of any doctoral programme.

Out in the real world, Carly and Moray have been wonderful flatmates and friends to me, even at my most grumpy. Ana (in cahoots with Hotel Chocolat) has kept me supplied with luxury chocolates on a student stipend– thanks just doesn’t cover it. And I may not have gotten through the last few years without Thomas, whose patience and generosity is astounding, and who has picked me up off the floor many times. Often literally.

Finally, I would like to thank my family for their continued moral and financial support. I would like to dedicate this thesis to my mum, who is at least partly responsible for the two decades of me leading up to this point, and who seemed to be somewhat surprised by the result. But she was the person who encouraged a love of learning that led me right here. I cannot give you a copy of this, so I’ll give you the whole thing instead.

Contents

List of Figures

List of Tables

Chapter 1

A cultural evolutionary framework for manual communication systems

1.1 Introduction

Human languages demonstrate an expressive power unrivalled amongst natural communication systems, that allow us as a species to practice an exceptionally wide range of behaviours, from simply attracting another's attention, to writing a doctoral thesis. Individual languages, whilst allowing the representation of largely similar world states, show substantial diversity, differing in structure, vocabulary and, in cases, modality (Evans & Levinson, 2009). The central endeavour of linguistic study is to explain why human languages look like they do and what these forms may tell us about our underlying cognitive or cultural preferences. Unsurprisingly then, the study of language evolution attempts to understand why human language looks the way it does, posing several key questions. How do humans learn the language they are exposed to, and how does this learning process reflect human cognition? How do cross-linguistic similarities and differences reflect the cognitive and social preferences of learners? And how did present linguistic forms develop over the evolutionary time scale of human languages?

This thesis is fundamentally a study of the evolution of human language, aiming to elucidate problems concerning how domain-general learning and interactive mechanisms drive the emergence of linguistic structure out of non-linguistic communication, over generations of linguistic populations. I build on previous work that proposes a cultural evolutionary account of language emergence, such that the systematic structure found in human languages is shaped by the cognitive and cultural biases of individuals and populations. The work described in this thesis provides an in-depth investigation of these domain general biases, to understand the role that individual mechanisms play in the evolution of language. In the course of this investigation, I demonstrate a novel methodology, artificial sign language learning, that can be used to further understand the full range of tools humans use to produce, use and process language. A lack of observable evidence makes studying the evolution of language difficult; experimental research allows the testing of participants preferences in a controlled environment. This research lays the foundations for closer ties between experimental research and field research, using natural language data from emerging sign languages to inform the experimental models detailed here.

Much of this body of work draws from relevant research into the manual modality: sign languages of the Deaf, and other gestural systems. The value of observing communicative capacities in the manual modality cannot be overstated, allowing the observation of language learning contexts that can differ from spoken language learning. Though deaf children with normal linguistic input acquire sign languages as easily as hearing children acquire spoken languages, many deaf children do not receive consistent linguistic input (Goldin-Meadow, Mylander & Butcher, 1995; Sandler, Aronoff, Meir & Padden, 2011). Individual deaf children deprived of normal linguistic input innovate gestural signals in order to communicate, creating homesign systems that are used to interact with their immediate family members (Goldin-Meadow, 2003; Goldin-Meadow, Butcher, Mylander & Dodge, 1994; Goldin-Meadow & Mylander, 1983). Communities of deaf individuals who do not have a linguistic model create new languages that can be observed as they evolve in real time (Sandler, Meir, Padden & Aronoff, 2005; A. Senghas, Kita & Ozyürek, 2004). This particular set of sign languages, emerging sign languages, have been the focus of much study in recent years, allowing the study of a language from its point of creation, as well as the processes that affect them, such as population size, social structure and age of the language (Meir, Assaf, Sandler, Padden & Aronoff, 2005). Evidence of this kind is unavailable in the spoken modality and provides crucial evidence of language creation and emergence. Moreover, the full range of human manual communicative abilities allows the study of communication in the individual (homesign), communication in a new community (emerging sign languages), and communication in a historically established community (older sign languages, such as American Sign Language), offering an overview of human language at several possible evolutionary stages. The commonalities between signed and spoken languages are indicative of the shared mechanisms that underlie both systems and thus serve to expand our understanding of general linguistic capacities, while differences between the two modalities point to modality-specific features that may only be evidenced in comparison (Sandler, 2013).

I explore both the modality-general and the modality-specific in this work. The aim of this thesis is to provide a precise understanding of the roles that the individual and the community play in the evolution of systematic linguistic structure; in particular, how interaction between individuals and transmission of a communication system to new individuals affects linguistic structure. Informed by what is currently the only available data from naturally emerging languages in new populations, emerging sign languages, the experiments I discuss seek to investigate how the cultural evolutionary mechanisms of interaction and transmission shape human communication systems, driving their evolution from restricted, unstructured systems to language-like systems that are both learnable and communicatively expressive.

1.1.1 Chapter outline

The remainder of this chapter will provide the background to the experimental work that makes up the main body of this thesis. I will introduce key concepts that define the main areas of study of this thesis, and discuss the experimental and theoretical research that motivates the present body of work. In section 1.2, I define the concept of systematicity in language, as the focus of my study, and assert its importance as a fundamental structuring principle of language. In section 1.3, I discuss language as a product of cultural evolution, viewing the study of systematicity from a cultural evolutionary perspective. I review previous experimental literature that has attempted to model learners by themselves, or in populations that learn from and communicate with each other. Section 1.4 highlights the importance of research into sign languages, as well as other manual communication systems, as a companion tool to experimental research in the cultural evolution of language. In section 1.5, I detail studies that have begun to bridge the gap between the experimental research described in 1.3 and the and the natural language research from section 1.4, before outlining the present methodology, artificial sign language learning, in section 1.6. I propose that artificial sign language learning is a substantial extension of previous experimental research, one that uses data from emerging sign languages to inform experimental models that place manual communication in a cultural evolutionary framework. Finally, I provide an overview of the remainder of the thesis, setting out the main aims of this work and highlighting its contribution to the wider field of language evolution research.

1.2 Investigating systematicity

1.2.1 Recognising systematicity

This thesis principally aims to investigate emergence of systematic structure in language. Systematicity is a term that has had many different interpretations, but given few explicit definitions in linguistic study. Therefore, it is necessary to clearly set out the structures I intend to investigate under the umbrella of systematicity.

With its most basic definition, systematicity refers to the notion of the *system*, a notion that rests at the heart of how we define the syntax of human languages: that languages are systems of related elements, recombined to form an inexhaustible set of expressions. We are able to understand almost any logical, grammatical utterance, even if we have not heard it before, as a result of having learnt both the units of our language, as well as how those units are combined, or how they relate to each other.

This recombination of linguistic units occurs across different levels of language. At the phonological level, finite sets of sounds are recombined to allow the expression of a far wider range of utterances. Above the level of phonology, the systematic re-use and recombination of units reflects similarities and differences in the meanings that signals refer to. For example, in English we systematically signal that an event has previously taken place with the past tense form *-ed*, as in the sentence *I finish**ed** writing my thesis*. If I re-use this affix in another sentence, *I smil**ed** all day*, it is understood that the affix signals the same part of that meaning, i.e., that this is an event that has occurred prior to the present moment. Furthermore, the differences between the forms of the verb stems (finish vs. smile) also signal a difference in meaning, allowing disambiguation. Systematic variation allows language learners to produce an inexhaustible amount of expressions from a finite, learnable set of elements.

1.2.2 Linguistic systematicity as a reflection of cognition

Earlier definitions of systematicity from cognitive science attempted to explain how systematicaticity could account not only for linguistic structure, but also for the structure of mental representations more generally (Fodor & Pylyshyn, 1988; Johnson, 2004; Matthews, 1997). Though these definitions provide some clarification of the term, they do not sufficiently capture its use in the present investigation. For instance, Fodor and Pylyshyn (1988, p.37) define systematicity as the knowledge that:

> ”the ability to produce/understand some sentences is *intrinsically* connected to the ability to produce/understand certain others.”

As an example, the authors suggest that any native speaker of English who knows how to say that *John loves the girl* will also be able to say that *the girl loves John*. Though this proposition is accurate, it does not adequately capture the full structural power of systematicity. Firstly, it describes systematicity as part of the process of the learner. The structure of a system is already in place, and it is up to the learner to become aware of that structure. I assert that systematicity is a property of the system itself rather than a property of the learner, and the study of systematicity rests on explaining when a set of communicative signals is or is not systematic. Secondly, Fodor and Pylyshyn (1988) assert that systematicity is reliant on constituent structure, though systematic structure is possible without linguistic constituents, as it is present across levels of language, not just at the level of constituent structure.

Though other definitions of systematicity (Johnson, 2004; Pullum & Scholz, 2007) differ from Fodor and Pylyshyn (1988), viewing it as a property of languages themselves and not of the learner, their definitions (particularly from the perspective of Pullum and Scholz (2007)) capture the phenomenon in terms of *intersubstitutability*, such that systematicity occurs when elements of similar kinds can be substituted for one another whilst retaining grammaticality. Such definitions quickly run into problems, as direct substitutions frequently produce ungrammatical strings (for example, *I finished my thesis* in contrast to **I smiled my thesis*). Indeed, Johnson (2004) points to the lack of intersubstitutability as evidence that language is not systematic, and natural languages contain strings and sub-strings that do not systematically conform to conventional structure. Following the example given above, though *finished* and *smiled* exhibit clear systematicity (and intersubstitutability to some extent), there are cases that do not conform to this rule, for example, if *finished* is contrasted with *slept*. However, these definitions remain restrictive. Though direct intersubstitutability is not always possible, languages exhibit system-wide structural properties that allow expressive power from a learnable set of elements.

1.2.3 Systematicity as a representational property

A second major perspective on systematicity positions the term as a representational property, in a theoretical framework alongside iconicity and arbitrariness (Dingemanse, Blasi, Lupyan, Christiansen & Monaghan, 2015), to understand the range of signal-meaning mappings in languages. Iconicity, the resemblance between a linguistic signal and its meaning, sits in opposition to arbitrariness, the "unpredictable mapping" of form and meaning (Dingemanse et al., 2015). Systematicity, as a predictable mapping that is not grounded in perception of the world, represents the middle ground between these two extremes of representation. For example, the systematic use of the phonoaesthemem *sl-*, connoting negative properties in words such as *slime* and *sloppy*, does not offer a real-world representation of these concepts, but does render their meanings more predictable across the linguistic system, so that there is a non-arbitrary relationship between form and meaning (Monaghan, Shillcock, Christiansen & Kirby, 2014). Both experimental research and data from natural languages indicate an interaction between iconicity and systematicity, such that as systematic structure becomes widespread over a set of signals, those signals become less iconic (Fay, Garrod, Roberts & Swoboda, 2010; Frishberg, 1975; Garrod, Fay, Lee, Oberlander & MacLeod, 2007; Theisen, Oberlander & Kirby, 2010; Theisen-White, Kirby & Oberlander, 2011; Verhoef, Kirby & de Boer, 2014). Furthermore, the ability to produce iconicity inhibits the emergence of systematic structure in communicative signals (G. Roberts, Lewandowski & Galantucci, 2015; Verhoef, Kirby & de Boer, 2015), evincing a potentially oppositional relationship.

There are several reasons why this framing of systematicity remains insufficient. Firstly, it remains embedded in a debate about different representations of meanings; systematicity as a structuring principle of language is ignored in such investigations, and systematic structure in the absence of meaning remains unaccounted for. Secondly, the oppositional relationship between systematicity and iconicity is not necessarily clear-cut. Iconicity persists even where systematicity is present; sign languages that rely heavily on iconic visual representations are still systematically organised (Frishberg, 1975; Klima & Bellugi, 1979; Perniss, Thompson & Vigliocco, 2010). Furthermore, iconic signals themselves can be systematically organised, leading to systems that demonstrate ”patterned iconicity” (Padden, Hwang, Lepic & Seegers, 2014; Padden et al., 2013). Patterned iconicity occurs when different representational strategies are consistently varied to denote a relationship, such as between lexical items. Iconic representations can make use of different salient features of real world objects, states, or events. For example, a gestural signal to denote a cat could highlight different physical features (for instance, whiskers or a tail) or behavioural features (such as grooming or scratching). Many sign languages demonstrate patterned iconicity, where iconic variants signal different categories, such as the difference between handling forms (where the handshape represents a hand manipulating the object) and instrument forms (where the handshape represents the object itself) (Padden et al., 2014). The relationship between systematicity and iconicity is complex; it is not clear to what extent it is a complementary or an oppositional relationship. As such, positioning systematicity on a representational scale between iconicity and arbitrariness fails to capture the complex relationship between systematic structure and the representation of meaning.

1.2.4 Systematicity as a fundamental structuring principle

The two previous sections provide different, but not necessarily conflicting accounts of systematicity. Both accounts postulate systematicity as an explanation for linguistic representation, either as a mental representation, or as a mapping between form and meaning. Although these definitions capture well the structure and representational applications of systematicity in language, they do not fully describe the extent to which systematicity is present in natural languages.

This thesis is concerned with understanding how systematic structure emerges in a linguistic system and how it evolves as a language is learnt and used within a population. Previous applications of the term systematicity have either focussed narrowly on systematicity as a representational device that only describes the relationship between form and meaning (Dingemanse et al., 2015; Monaghan et al., 2014), or have relied on pre-existing structures (namely constituents) to scaffold definitions of systematicity (Fodor & Pylyshyn, 1988; Johnson, 2004; Pullum & Scholz, 2007).

To define systematicity here, I return to the key notion of the *system*, as a finite set of related elements that can be re-used and re-combined to provide a far wider range of expression. I propose that systematicity is a fundamental structuring principle of language, that can account for the structure found across different levels of linguistic enquiry. Finite, learnable sets of phonological, morphological, lexical and syntactic units are re-used and recombined in languages to create conventional, expressive signals. This definition may account for signal-meaning mappings; systematic re-use and recombination can denote relationships between the meanings that signals map onto (Kirby, Cornish & Smith, 2008), but it is not always the case (for example, phonology is systematic, but not with respect to meaning). As an umbrella term, systematicity therefore accounts for different types of linguistic structure, such as combinatoriality (re-use and re-combination without reference to meaning), compositionality (re-use and recombination with respect to meaning) and regularity (the consistency of form across meanings), and provides a framework by which different types of structure can be studied with the same theoretical and methodological tools.

The research conducted in this thesis exemplifies this framework, aiming to elucidate how cultural evolutionary mechanisms lead to the emergence and evolution of systematicity in different linguistic contexts, and how the cultural evolution of systematic structure is affected by representational devices such as iconicity. In the following section, I present a discussion of the cultural evolution of language, and how experimental models of language evolution provide a vital tool for understanding the emergence of systematic linguistic structure.

1.3 The cultural evolution of language

1.3.1 Language as a cultural product

Like language, culture is a term that has been defined in many different ways. Here, following Mesoudi (2015), the term 'cultural' accounts for socially transmitted behaviours or information, and thus covers a diverse and complex range of human social practices. Though the study of human culture is primarily concerned with behaviours, states and information, the similarities that can be drawn between cultural changes and biological changes suggest comparable underlying processes. These similarities underlie theories of cultural evolution, which assert that socially transmitted behaviours can be described and analysed as an evolutionary process, using a conceptual framework similar to that implemented in the study of biological evolution. In this framework, cultural variants are described in terms of their variation, their fitness (or rate of survival) and their inheritance between members of a population (Mesoudi, 2015; Mesoudi & Whiten, 2008), comparable with descriptions of biological variants. In other words, cultural products demonstrate variation, and different variants can be adopted to different extents, through the process of inheritance within and across generations of human populations.

Language, in many respects, should be the poster child for cultural evolution. Variation and diversity are present both within languages and cross-linguistically (Evans & Levinson, 2009), and variants do not co-exist equally, but are present in languages to different extents at different points in time (Blythe & Croft, 2016; Stadler, Blythe, Smith & Kirby, 2016). Finally, languages as acquired by children, are learnt through a transmission process, from generation to generation, though language change also operates within a generation, propagating through interaction between language users.

Viewing language as an evolutionary process has been applied successfully to studies of ongoing language change (Baxter, Blythe, Croft & McKane, 2006; Blythe & Croft, 2016; Croft, 2000; Stadler et al., 2016), but has also been applied to the study the evolution of language from non-linguistic states (Tamariz & Kirby, 2016). Such accounts suggest that the same processes that operate on modern languages (particularly the processes of interaction between language-users and transmission of a language to new learners) are stable processes that have operated throughout the evolutionary history of modern biological language capacities (meaning here the set of physiological and cognitive capacities that allow humans to learn and use language). This cultural evolutionary account of language necessarily proposes a uniformity of process throughout linguistic evolution (Motamedi, Schouwstra & Kirby, in press), which thus posits a continuity between non-linguistic communication and structured, complex linguistic systems. It is this transition, from non-linguistic communication to language that is the primary concern of cultural accounts of language evolution, as opposed to the biological evolution of the capacities to learn and use language (Tamariz & Kirby, 2016). The cultural evolution of language operates within populations of learners, in which individuals use the language to interact with each other, and in which the language is transmitted over time to new generations of learners. Thus, it is possible to highlight three key processes in the evolution of linguistic systems: learning in the individual, who have no learning model but produce linguistic behaviours that reflect their cognitive preferences, interaction between learners in a population, and the transmission of the system to new learners of the language. Investigation into the preferences of individual learners provides evidence of the cognitive biases that underlie particular linguistic structures, and are indicative of the functional pressures that shape language (Culbertson, Smolensky & Legendre, 2012; Hudson Kam & Newport, 2005; Saffran, Newport & Aslin, 1996). Interaction between language users is part of the key communicative function of language; as such, the pressures that interaction applies to language use could serve to further shape the types of structures we see in language. Finally languages are acquired through learning. New language learners (human infants) are born into the linguistic community and must learn the language of that community from observing existing language users. Studying transmission between language learners provides evidence of how systems change over generations of learners in a linguistic community.

Furthermore, these processes are not independent of each other. Individual learners make up populations that interact and learn from each other. Interaction does not entail that no learning takes place, as suggested by the many linguistic changes that are propagated through populations by mature language learners. Generational transmission involves (and may necessitate) interaction between infant and carer; children exposed to linguistic input but deprived of interaction do not acquire full linguistic competence (Hoff, 2006). Therefore, understanding the role of each process is crucial for understanding how these processes combine over the evolution of a language to develop the structures found in natural human languages. Below, I describe the wide body of experimental research that has attempted to understand these processes in turn, and in relation to each other.

1.3.2 Modelling the cultural evolution of language

Artificial language learning in the individual

Artificial language learning paradigms expose participants to artificial languages, training them on a system, or part of a system, before subsequently testing them on their ability to either reproduce the language (Clay, Pople, Hood & Kita, 2014; Culbertson et al., 2012; Hudson Kam, 2015; Hudson Kam & Newport, 2005) or to recognise structures from the language (Culbertson & Adger, 2014; Saffran et al., 1996; St. Clair, Monaghan & Ramscar, 2009). Experimental research implementing artificial language learning has demonstrated evidence for differences in how children and adults learn (Clay et al., 2014; Hudson Kam, 2015; Hudson Kam & Newport, 2005), for statistical learning capacities of language learners (Saffran et al., 1996), and for particular structural preferences that reflect typological preferences cross-linguistically (Culbertson & Adger, 2014; Culbertson et al., 2012; St. Clair et al., 2009). These paradigms have been applied across a range of experimental contexts to investigate the underlying cognitive preferences that individual learners exhibit through linguistic behaviour. Furthermore, artificial language learning experiments on individual participants have produced findings suggesting that learners adapt language for communication, even without the presence of a partner (Fedzechkina, Jaeger & Newport, 2012). The application of these paradigms highlights the role of the learner and individual cognition in shaping linguistic structure, and, as the results from Fedzechkina et al. (2012) suggest, indicate how individual learners adapt language for its communicative function. However, individuals do not develop language in isolation but learn languages as part of communities that interact with and learn from each other. Understanding only the role of the individual does not provide a sufficient account of how or why the structures found across human languages emerge.

Experimental models of transmission

Experimental models of transmission have used artificial language learning paradigms, implementing these methods into a cultural evolutionary framework where participants learn an artificial language transmitted to them by another participant, to model the evolution of communication systems in populations of individuals. Iterated learning refers to the model of transmission that is applied across the majority of these experiments, denoting the process whereby one individual learns a behaviour, via observational learning, from other members of the same population who have previously learnt that behaviour in the same way (Kirby, Griffiths & Smith, 2014; Smith, Kirby & Brighton, 2003; Tamariz & Kirby, 2016). Or, simply, the behavioural output produced from one member of the population becomes the learning input for a new member of that population. This is analogous to the experience of the infant learner, who acquires the language of their community through observation of the linguistic output of their parents or other community members, who themselves learnt the language through observation of other members of their linguistic community.

The application of the iterated learning model in experimental research has demonstrated the emergence of systematic, language-like structures from randomly generated signals, through the process of learning (Kirby et al., 2008). Kirby et al. (2008) presented participants with randomly generated signals in a training stage and then asked participants to reproduce those signals in testing. The labels they produced in testing were subsequently used as training for another participant, forming a transmission chain where each individual participant represented a generation in the population. As the sets of labels were learnt at each generation, the accumulation of changes to the signals through learning led to structured labels that reflected the dimensions of the meaning space: in the case of Kirby et al. (2008), colour, shape and manner of motion. The systems produced by participants reflected the pressures applied by the transmission process, leading to structures that were simpler and easier to learn.

Further expansion of experimental research into the cultural transmission of language has complemented earlier results, showing that the transmission process facilitates the emergence of systematic structures, both compositional (Cornish, Smith & Kirby, 2013; Kirby et al., 2008; Kirby, Tamariz, Cornish & Smith, 2015) and combinatorial (Verhoef et al., 2014, 2015), as well as facilitating the emergence of regularity (Smith & Wonnacott, 2010). Experimental work in non-linguistic domains has highlighted the domain-general nature of the learning process that leads to the emergence of structure in language, demonstrating similar results in musical tasks (Ravignani, Delgado & Kirby, 2016; Verhoef et al., 2014, 2015) and in a non-linguistic sequence-learning task (Cornish et al., 2013). Moreover, experiments that apply the iteration process across different linguistic domains suggest that cultural transmission is responsible for both systematic sequential ordering (Kirby et al., 2008, 2015; Winters, Kirby & Smith, 2015) and the emergence of category structures in language (Carr, Smith, Cornish & Kirby, 2016; Carstensen, Xu, Smith & Regier, 2015; Silvey, Kirby & Smith, 2014; Xu, Dowman & Griffiths, 2013), as well as the emergence of symbolic signals that are learnable by naive users (Caldwell & Smith, 2012). Though many of these studies ask participants to communicate concepts in minimal, highly constrained meaning spaces, the structuring of signals according to salient features across meanings has been shown in more complex, continuous meaning spaces (Carr et al., 2016; Perfors & Navarro, 2014). Furthermore, studies that model the transmission of communicative signals have been used to understand the interaction between linguistic structure and iconic representation (Verhoef et al., 2015), suggesting that the ability to represent meanings iconically inhibits the development of linguistic structure.

This wide body of experimental research demonstrates how populations of learners affect the forms found in a linguistic system, across different linguistic domains, and across different referential contexts, highlighting the importance of domain-general learning processes in shaping language (Christiansen & Chater, 2008; Culbertson & Kirby, 2016). The implication that the transmission of a system across learners significantly shapes linguistic structure substantially departs from earlier frameworks of language evolution that suggest the structures found in language reflect underlying domain-specific biases that were selected for during biological natural selection (Pinker & Bloom, 2000).

Experimental models of interaction

Concurrent with the development of experimental work on generational transmission, interaction research privileged the inherently social aspect of human language: that it is used as a tool for communication. Early research into the effects of interaction on verbal communication between pairs of participants pointed to the collaborative role of communication in the development of signals that both enable comprehension between interlocutors and minimise production effort (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Krauss & Weinheimer, 1966).

The expansion of the experimental interactive framework to non-linguistic tasks further highlights interaction as a fundamental, domain-general process that shapes communication more generally outside of the linguistic domain. The graphical communication paradigm, in which participants communicate concepts by drawing them (similar to the board game Pictionary), demonstrates the importance of interaction as a mechanism for shaping linguistic structure (Fay, Arbib & Garrod, 2013; Fay et al., 2010; Garrod et al., 2007; Healey, Swoboda, Umata & King, 2007). Participants across these studies were instructed to communicate a set of concepts to a partner, using drawings, over several rounds, repeatedly drawing the same concepts. As shared knowledge between participants was established, drawings became increasingly symbolic and less complex to produce, balancing the need for comprehension between communicators and minimised production effort. Fay et al. (2010) and Fay and Ellison (2013) contrasted pairs of communicators and communities of interacting individuals that had to negotiate a shared system in groups. The same pressures for accurate but efficient communication operated over communities of individuals, producing symbolic sets of signals. However, convergence on a shared system occurred more quickly in pairs than in larger communities (Fay et al., 2010), though Fay and Ellison (2013) found that the benefits of increased identification accuracy and reproduction fidelity in the community context could override the benefits of speedy convergence. The findings across these graphical communication experiments suggest that communication is an alignment process, where interlocutors must negotiate common ground in order to converge on a shared system that allows participants to maximise comprehension whilst minimising production effort. The signals participants produce reflect this alignment process; they become increasingly symbolic, reflecting the esoteric knowledge shared by the pair, or group, of participants.

In addition to the graphical communication studies, research in the field of experimental semiotics has modelled the emergence of communicative signals through interaction when there is little possibility for any initial shared knowledge (Galantucci & Garrod, 2011). Galantucci (2005) presented a signalling game where pairs of participants had to navigate through space in a video game to achieve a common goal. However, their representation of that space was limited, such that they could only see the 'room' in which their avatar was located at a particular time, and were otherwise blind to the layout of the space they had to navigate. Furthermore, conventional signals were inhibited through the medium of communication, a digitizing pad that did not allow rendering of letters, numbers or shapes. Participants were still able to develop successful signalling systems, relying on iconic representations such as lines that represented numeric values, or that represented positions in space. Over rounds of communication, participants built common ground through completing the task, and were thus still able to negotiate a shared signalling system. G. Roberts and Galantucci (2012) presented an experimental semiotics study in which participants had to communicate a set of concepts, but as in Galantucci (2005), conventional signals were inhibited via the digitizing pad that was used to send signals. Participants produced signals that showed the recombination of a small number of unique forms over rounds in the experiment, exhibiting properties comparable with combinatoriality in natural languages. In a later study, G. Roberts et al. (2015) implemented a similar methodology to investigate the interference that iconicity has on the development of combinatoriality. Participants were either assigned to a mimeable condition, where the concepts in the meaning space could be represented iconically, or to a non-mimeable condition, in which iconic representations were not possible. Participants in the non-iconic condition produced combinatorial signals over rounds of communication, replicating G. Roberts and Galantucci (2012); participants in the mimeable condition produced signals that did not demonstrate combinatoriality. The ability to represent signals iconically inhibited the development of structured signals, suggesting that pressures from communication can interact with the representational mechanisms that shape language.

Finally, Scott-Phillips, Kirby and Ritchie (2009) demonstrate the emergence of communicative signals where overt signalling is not immediately available to participants. Pairs of participants who had to navigate through a grid of four coloured squares negotiated a communication system that facilitated this navigation, though they were given no signalling medium with which to communicate other than the game itself. Participants used movement of their avatars within the space to communicate aspects of the space to their partners, negotiating a shared signalling system based on their shared knowledge of the space they had to navigate. Through the development of this shared knowledge over rounds of interaction, participants created successful signalling systems that were adapted to the task they were required to complete.

As with investigations into the transmission of communicative signals, the experimental research into the effects of interaction suggest that interaction is a driving force in shaping linguistic systems. Participants with no prior history develop shared knowledge over rounds of communication, using that shared knowledge to negotiate a system that allows mutual comprehension, at the same time as minimising the collaborative effort of the communicators.

Combining transmission and interaction

Although the research described above offers a vital insight into the effects of both transmission and interaction, the separation of transmission and interaction in these cases presents a problem. Both interaction and transmission are at work in natural languages; we both learn from other generations of language learners and communicate with those within our linguistic community. The experimental research detailed above suggests that both interaction and transmission lead to similar linguistic structures (G. Roberts et al., 2015; Verhoef et al., 2014, 2015), although without observing each mechanism at work in a comparable task, it is difficult to remove other factors that may be at work (such as the effects of iconicity evidenced by both G. Roberts et al. (2015) and Verhoef et al. (2015)). As such, to fully understand the effects that transmission and interaction have on the forms of language found in the world, it is necessary to understand the roles that each mechanism plays in relation to the other.

Recent experimental work has begun to investigate the precise effects that interaction and transmission have on linguistic structure. Kirby et al. (2015) detail an experiment where pairs of participants had to communicate a meaning space to each other that consisted of different shapes distinguished by different patterns and appendages. Participants took turns in a signalling task in which one participant in a pair produced a label for a meaning, and the other participant interpreted the label by selecting a meaning from the meaning space. To compare the effects of interaction and transmission in combination against the effects of interaction alone, the authors contrasted chains of pairs of participants (in which participants interacted, and then the signals they produced were transmitted as training to a new pair of participants), with closed pairs of participants (who repeatedly communicated with each other without the introduction of new learners). Their findings suggest that communicative pressure alone, present in the closed pair condition, leads to holistic signals, idiosyncratic to a particular pair, but that the combination of interaction and transmission leads to signals that demonstrate systematic structure, with participants recombining parts of signals across the meaning space to signal similarities between meanings. Other studies investigating the combination of interactive and transmissive pressures have suggested that both pressures together lead to a cumulative rise in systematicity over a set of signals, in comparison to interaction alone. Theisen et al. (2010) and Theisen-White et al. (2011) contrast interaction alone in a graphical communication task with the combination of interaction and transmission. In both studies, pairs of participants took part in a task similar to that detailed by Garrod et al. (2007) and Fay et al. (2010), communicating concepts to a partner by drawing the concepts, for their partner to try and interpret. The concepts participants had to communicate shared semantic associations, either on a thematic dimension (such as fireman and fire engine) or on a dimension termed 'entity type' (for example, fireman and teacher sharing an association, both referring to the profession a person may belong to). Theisen et al. (2010) replicated previous experimental results showing that interaction leads to more symbolic signals; however, they also demonstrated that systematicity emerges in these signals, but does not accumulate over time, as participants communicate. In contrast, Theisen-White et al. (2011) showed that the addition of transmission pressures led to a cumulative increase in systematicity over generations of communicating pairs. Furthermore, these results were further suggestive of the interaction between iconicity and systematic structure; as systematicity increases, signals become less iconic. Carr et al. (2016) also demonstrated that the combination of interaction and transmission led to increased systematicity in comparison to interaction alone, in a more complex meaning space. Pairs of participants in transmission chains had to communicate about a set of meanings that represented triangles, which could differ across a number of parameters, such as size, location, and rotation, presenting a continuous and expansive meaning space for participant to disambiguate. Participants in the combined condition (interaction and transmission together) produced signals that demonstrated the evolution of sub-lexical category structures over time. In comparison, interaction alone led to a bigger range of expressive category labels that did not systematically recombine parts of signals. Finally, a third experiment investigating the effect of transmission alone found the emergence of small sets of learnable category labels that lacked expressivity.

Although the above results indicate that interaction and transmission respond to different pressures and, as such, lead to the emergence of different structures, other experimental work contrasting the two mechanisms has suggested that they can also lead to the emergence of similar structures. Schouwstra, Smith and Kirby (2016) presented a study that contrasted communicative and transmissive pressures in a gesture task. Participants asked to communicate events to a partner using only gesture produced gestures initially conditioned on the types of events they had to gesture. In both an interaction-only condition and a condition that combined interaction and transmission, semantically conditioned word order was replaced by regular word order at a comparable rate. These results, in contrast to the other studies discussed in the present section, suggest that regularity in language (the consistent use of form across contexts) can emerge without transmission to new learners, and demonstrate how different types of systematicity may emerge in response to different pressures.

There is a wide body of experimental research described in this section, which has attempted to understand the effects that individual cognition, interaction between individuals and transmission of language between individuals have on the signals, and the systems of signals, that participants produce (see table 1.1 for summary of the works cited in this section [1]). However, the precise effects of different cultural evolutionary mechanisms in comparable contexts are by no means clear and uncontested. Furthermore, the need to link experimental research back to natural languages is evident - the experimental models we use must be able to account for the data that we see in the phenomenon of interest, human language. In section 1.4, I discuss evidence from the evolution of manual communication systems and sign languages that provide a valuable companion to experimental language evolution research.

[1] In their review of cultural evolution experiments, Tamariz and Kirby (2016) also provide a summary table which groups experiments by task and input that can be consulted for further clarity.

Individual production	
Developmental differences between children and adults	Clay et al. (2014); Hudson Kam (2015); Hudson Kam and Newport (2005)
Statistical learning capacities in adults	Saffran et al. (1996)
Structural preferences reflect typological tendencies	Culbertson and Adger (2014); Culbertson et al. (2012); St. Clair et al. (2009)
Individuals adapt signals for a communicative function	Fedzechkina et al. (2012)
Transmission alone	
Evolution of compositionality	Cornish et al. (2013); Kirby et al. (2008, 2015); Winters et al. (2015)
Evolution of combinatoriality	Verhoef et al. (2014, 2015)
Evolution of regularity	Smith and Wonnacott (2010)
Evolution of categorical structure	Carr et al. (2016); Carstensen et al. (2015); Perfors and Navarro (2014); Silvey et al. (2014); Xu et al. (2013)
Evolution of symbolic representation	Caldwell and Smith (2012)
Interaction alone	
Emergence of collaborative communication	Brennan and Clark (1996); Clark and Wilkes-Gibbs (1986); Galantucci (2005); Krauss and Weinheimer (1966); Scott-Phillips et al. (2009)
Evolution of symbolic representation	Fay et al. (2013); Fay and Ellison (2013); Fay et al. (2010); Garrod et al. (2007); Healey et al. (2007); Theisen et al. (2010)
Evolution of combinatoriality	G. Roberts and Galantucci (2012); G. Roberts et al. (2015)
Evolution of regularity	Schouwstra et al. (2016)
Interaction and transmission (combined)	
Evolution of compositionality	Kirby et al. (2015); Theisen-White et al. (2011)
Evolution of sub-lexical structure	Carr et al. (2016)
Evolution of regularity	Schouwstra et al. (2016)

Table 1.1: Summary of experimental work reviewed in section 1.3, organised by cultural mechanism, as given in sub-section headings. Cited works under each mechanism are grouped according to principal outcome of the experimental work, and the structural properties exhibited by signals participants produced.

1.4 Insights from sign language research

The evolution of language has, for much of the history of the field, been synonymous with the evolution of speech, and modern human sign language capacities have not been the focus of study. This slow integration of sign language linguistics in theories of language evolution is perhaps down to the late acceptance of sign languages as fully-fledged languages, following efforts to document American Sign Language (ASL) in the 1960s (Klima & Bellugi, 1979; Stokoe, 1960). Until more recently, discussion of the manual modality in language evolution research has focussed on the possibility of its role earlier in evolutionary time, as a gestural protolanguage, a precursor to modern human language (Arbib, Liebal & Pika, 2008; Corballis, 2003, 2010; Kendon, 2016; Sterelny, 2012). Research indicating greater flexibility and productivity of non-human primate gestures in comparison to their vocalisations (Gillespie-Lynch, Greenfield, Feng, Savage-Rumbaugh & Lyn, 2013; Hobaiter & Byrne, 2014; A. Roberts, Vick, Roberts & Menzel, 2014; Slocombe, 2012), as well as greater success in teaching sign languages to non-human primates (Bronowski & Bellugi, 1970; Miles, 1990; Patterson, 1978), has led to various theories that posit manual communication as the dominant modality for communication in our pre-linguistic hominid ancestors.

However, the main aim of this thesis is to understand the evolution of linguistic structure once a communicative platform (whether gestural or vocal) is already established. More recent integration of sign language research into studies of evolutionary linguistics as a tool for understanding the development of modern linguistic structure (i.e., rather than viewing gesture as a more primitive stage of human communicative capacities) has provided a wider understanding of both modality-independent properties of language, as well as modality-specific phenomena (Meier, 2004; Sandler, 2013; Sandler & Lillo-Martin, 2006). The variation in exposure to a linguistic model that deaf children face leads to widely differing systems depending on whether those children are part of a linguistic community or isolated from normal linguistic input, and thus allows the observation of communication systems at different stages of evolution. In particular, the manual modality provides the only current accessible natural language evidence of the birth of new languages, emerging sign languages that are being documented in several different communities globally (de Vos, 2014; H. E. Morgan, 2015; Nyst, 2010; Sandler et al., 2005; A. Senghas et al., 2004). For the field of language evolution, which crucially lacks observable evidence of the object of study, these newly created languages offer insight into the first stages of linguistic creation.

1.4.1 Established sign languages

Under the term established sign languages, I count documented sign languages that are known to be over a 100 years old, that are used within a large community and that are often considered the standard language of the Deaf within a country. For example, American Sign Language (ASL), British Sign Language (BSL), Australian Sign Language (Auslan), sign language of the Netherlands (NGT) and German Sign Language (DGS) fall under the umbrella of this term. I will contrast these sign languages with homesign systems in individuals and family units, and with emerging sign languages, that are less than 100 years old. Research into ASL in the 1960s and 1970s, particularly the work of Stokoe (1960) and Klima and Bellugi (1979), led to wider acknowledgement of sign languages as fully-structured linguistic systems, and ASL remains one of the most well-documented sign languages.

Linguistic study of sign languages has revealed significant structural similarities between signed and spoken languages; sign languages demonstrate phonological, morphological and syntactic structures comparable with spoken languages (Klima & Bellugi, 1979; Meier, 2004; Sandler & Lillo-Martin, 2006; Stokoe, 1960; Supalla & Newport, 1978). They can express different registers depending on social context, and show a rich history of creative expression through poetry (Klima & Bellugi, 1979; Sutton-Spence & Woll, 1999). The study of sign languages can evidence those structures and linguistic processes that are independent of modality, and offer a greater understanding of the human capacity for language. However, study of sign languages has also highlighted wide ranging differences in linguistic structure that point to modality-specific phenomena shaping the linguistic system (Meier, 2004; Sandler, 2013). As systems that rely on the visual representation of referents, sign languages exhibit high levels of iconicity, representations that are not afforded by the vocal modality (Perniss et al., 2010, though see Perlman and Cain (2014) for a discussion that contests this). The affordances of iconicity lead to very different linguistic structures, such as representations that are grounded in the signer's body (Taub, 2001), or grounded in the space around the signer (Emmorey, Tversky & Taylor, 2000; Liddell, 2003; Lillo-Martin & Meier, 2011). Sign languages are able to combine units simultaneously (Aronoff, Meir & Sandler, 2005; Sandler & Lillo-Martin, 2006; A. Senghas et al., 2004), in contrast to the sequential concatenation of linguistic units that characterises linguistic structure in spoken languages. Sandler (2013) asserts that understanding these differences between signed and spoken languages is essential for understanding the flexibility and adaptive ability of human cognition across both modalities.

In addition to the structural differences between signed and spoken languages, the acquisition context in which many deaf children learn sign languages differs from spoken language acquisition. Deaf children exposed from infancy to sign language acquire the language much like a hearing child acquires spoken language. However, current estimates suggest that more than 90% of deaf children are born to hearing parents (Goldin-Meadow et al., 1995; Sandler et al., 2011), most of whom are not users of a sign language. As such, a relatively high proportion of deaf children do not receive exposure to a consistent linguistic model in early childhood. Studies of late acquirers of sign languages show the difficulties deaf adolescents have in successfully acquiring a linguistic system, providing support for theories of a critical period of language acquisition (Mayberry, 2010; Mayberry, Lock & Kazmi, 2002). Furthermore, Singleton and Newport (2004) used the unique acquisition context of Simon, a young child who was learning ASL from late-acquirer parents, to demonstrate the regularisation capacities of the child learner. Simon, acquiring the system within the proposed critical period, regularised inaccurate ASL input from his late-acquirer parents, producing a linguistic system that showed greater consistency with standard ASL than the inconsistent input.

Generally, linguistic study of sign languages has had a great impact on our understanding of the human language capacity. For investigations into the emergence of linguistic structure in particular, the study of manual communication systems offers a unique view of what a linguistic system looks like at the earliest stages of its evolution, and how mechanisms such as the use of a system in interaction and learning contexts can lead to drastic changes in linguistic structure. The isolation of many deaf individuals from a consistent language model allows the uniquely informative observations of the spontaneous emergence of communication systems without prior linguistic knowledge, both in individuals and in newly established communities. I discuss the relevance of both of these contexts to the study of language evolution below.

1.4.2 The role of the individual: homesign systems

Homesign systems are systems of gestures developed by deaf children who do not receive conventional linguistic input (Goldin-Meadow, 2003; Morford & Goldin-Meadow, 1997). These children develop gestures to communicate with those around them, usually hearing family members. Homesigning children develop a stable lexicon, made up of primarily iconic signs and pointing gestures, and these systems have been shown to take on some of the characteristics of developed languages, such as basic word categories (Goldin-Meadow et al., 1994; Hunsicker & Goldin-Meadow, 2013), grammatical roles (Coppola & Newport, 2005), displaced reference (Morford

& Goldin-Meadow, 1997) and relational marking (Goldin-Meadow & Feldman, 1977), as well as some consistency in basic sign order (Goldin-Meadow, 2003). Nonetheless, they are not languages (Hoff, 2006; Slobin, 2004). They lack a community, usually only being used as a primary system by the child creator; often the gestural systems are not transmitted, but live and die within a single generation. Though homesigning appears to be a natural impulse in the absence of conventional linguistic input, systematic study of homesign did not occur until the 1970s, when the phenomenon began to be documented more closely. Since then, homesign systems have been studied in both urban, industrialised communities (Goldin-Meadow, 2003; Goldin-Meadow & Mylander, 1998), as well as in rural communities around the world (Haviland, 2013; Horton, Goldin-Meadow, Coppola, Senghas & Brentari, 2015; Nyst, 2010). Homesigns usually occur in family settings where a single deaf child lacks access to spoken language, but has also been documented in families with multiple deaf siblings (Haviland, 2013), though the community does not extend beyond the family.

Homesign systems evidence the spontaneous creation of a communicative system in a child deprived of linguistic input. The structures that are consistently present in these restricted systems can be viewed as evidence of basic cognitive preferences for structuring information, or of the fundamental properties of human language, the "resilient" properties of language (Goldin-Meadow, 2003), and may also offer insight into the development of other cognitive capacities, such as arithmetical abilities (Spaepen, Coppola, Flaherty, Spelke & Goldin-Meadow, 2013). Homesign systems can also suggest the extent to which the communicative function of these signals affect the form of the system. For instance, the structures found in homesigns do not appear to be affected by parental input (Goldin-Meadow & Mylander, 1983) and are not always fully comprehensible by the immediate family members that homesigners most frequently communicate with (Carrigan & Coppola, 2017). However, convergence with hearing communication partners does increase over time, as the system is used in interaction (Richie, Yang & Coppola, 2014).

Homesign has previously been suggested as providing a valuable window on to language evolution through the observation of linguistic creation in individuals, particularly in contrast with the learning of fully developed languages (Botha, 2007; Schouwstra, 2012). However, as Slobin (2004) and Botha (2007) have acknowledged, the insights gained from observation of homesign systems is limited; they are not full languages, nor do they develop in an altogether natural social context. Instead, homesigns represent one-sided systems in which a primary user communicates with interlocutors who are not only not fluent in the system, but who supplement communication in a modality inaccessible to the homesigner. Homesign can serve to bring to light common cognitive preferences that may underlie structures found in more developed linguistic systems; they point to the role of the individual learner in shaping communication systems. However, individual homesign systems do not develop into full languages (A. Senghas, 2005; Slobin, 2004). A community of communicators is required for a linguistic system to develop, highlighting the role of cultural evolutionary factors in the evolution of language. A further gain from studying homesign is in comparison with emerging linguistic systems within communities, as well as with established sign languages, to understand how adaptive pressures that the individual brings to bear on linguistic structure differ from those brought about through the use and transmission of a linguistic system. Observing the development of homesign systems provides a unique insight into linguistic creation in the individual, without a contributing community. Emerging sign languages evince the next stage of language creation after homesign: the establishment of a linguistic community and the early emergence of a language.

1.4.3 Language creation in a new community: emerging sign languages

Emerging sign languages are sign systems that have emerged over the last century, approximately, and thus allow the observation of a linguistic system from its creation. These new sign languages emerge following the establishment of a new Deaf community, such as through changes in educational policies that bring previously isolated deaf individuals together (H. E. Morgan, 2015; H. E. Morgan & Mayberry, 2012; A. Senghas et al., 2004). Studying emerging sign language provides a window onto language from its creation that is not found elsewhere in natural language environments; pidgins and creoles, which represent the only instance of novel spoken languages, are based on (and influenced by) existing source languages (Meir, Sandler, Padden & Aronoff, 2005). As with homesign, emerging sign languages evince the structures and representations that emerge early in a linguistic system, and point to those which require further evolution to emerge. Unlike homesign systems, however, the presence of a community and the further expansion of that community allows researchers to track particular linguistic properties, and how these properties evolve as the language is used and transmitted.

Research into emerging sign languages categorises these systems according to the social context in which they arise, into either *community sign languages* or *village sign languages* (Meir, Assaf et al., 2005; Meir, Sandler et al., 2005). Below, I outline the main features of each category, and discuss the impact that data collected from both types has had on language evolution research.

Community sign languages

Community sign languages arise when groups of deaf individuals are brought together through changes in educational policy, such as the establishment of schools for the deaf, thus forming a new Deaf community. The individuals who form these communities are typically from different regions within a larger area, and may be from different social or cultural backgrounds. They may have some prior linguistic knowledge, as is the case with Israeli Sign Language (Meir, Padden, Aronoff & Sandler, 2007; Meir, Sandler et al., 2005; Tkachman & Sandler, 2013), or individuals in the community may have established their own individual homesign systems, that they previously used to communicate with family members.

The latter situation provided the context of emergence for one of the most well-documented emerging sign languages, Nicaraguan Sign Language (henceforth, NSL: Goldin-Meadow, Brentari, Coppola, Horton & Senghas, 2014; A. Senghas, Coppola, Newport & Supalla, 1997; A. Senghas et al., 2004; R. J. Senghas, Senghas & Pyers, 2005). Nicaraguan Sign Language emerged in the Nicaraguan capital, Managua, in the late 1970s, following the establishment of a private school for the deaf, which later became public due to changes in health and literacy policy (A. Senghas et al., 1997). The children, many of whom used their own homesigns within their families, began communicating with each other using innovated gestures. Since then, research into NSL has recognised three main cohorts of students into the school, and the delineation of these cohorts can be used to trace how signs develop over time. For example, NSL showed the early development of argument structure and spatial reference (A. Senghas, 2000; A. Senghas & Coppola, 2001; A. Senghas et al., 1997), and has demonstrated the gradual evolution of linguistic structure over cohorts of the system, such as emerging lexical categories (Abner, Flaherty, Stangl, Brentari & Goldin-Meadow, 2016; Goldin-Meadow et al., 2014), temporal language (Kocab, Senghas & Snedeker, 2016a), phonological structure (Brentari, Coppola, Cho & Senghas, 2016), grammatical number (Horton et al., 2015), and recursive structures (Kocab, Senghas & Snedeker, 2016b). Studying the emergence of structure in NSL allows the investigation of the impact that the linguistic community (and the expansion of that community over time) has on the structure of the language. In particular, the school context has provided fertile ground for the emergence of many established sign languages, such as ASL and BSL, and as such offers a window onto the creation of a typically emerging urban sign language.

Israeli Sign Language (ISL) provides a slightly different context of the evolution of a community sign language. Emerging in the 1930s, ISL brought together signers from a range of different geographical and cultural backgrounds, some of whom had prior knowledge of a sign language (Meir et al., 2007; Meir, Sandler et al., 2005). Currently, the Israeli signing community numbers around 10,000 members, that can be grouped into four generations of signers. Over 80 years and these four generations, ISL has exhibited the evolution of a consistent phonology (Sandler et al., 2011), morphology (Aronoff et al., 2005), grammatical categories (Tkachman & Sandler, 2013), classifier constructions (Aronoff, Meir, Padden & Sandler, 2003), and the beginnings of a spatial agreement system (Meir et al., 2007; Padden, Meir, Aronoff & Sandler, 2010). Consistencies across emerging sign languages such as ISL and NSL can provide a window onto early emerging properties of language, but these differences can also offer an exploration of how the social structure of a community can affect the linguistic structures that emerge. These differences are even more pronounced between community sign languages and village sign languages, and serve to highlight the role of cultural mechanisms in shaping language.

Village sign languages

Village sign languages emerge in small, generally rural communities that have high rates of hereditary deafness. The proportion of deaf people in these communities is higher than in typical human communities (Meir, Sandler et al., 2005), and as such, a deaf community is established in which a sign language can emerge. The communities themselves are often isolated from the larger regional communities, for reasons of geographical distance, or ethnic or cultural differences, though within the community, deaf individuals are typically integrated normally into the community (Zeshan, 2007), and sign languages that emerge in these communities are often used by deaf and hearing individuals alike. Though generally less well-documented than urban or community sign languages, instances of emergent village sign languages have been documented across different geographical and cultural contexts throughout the world, such as Al-Sayyid Bedouin Sign Language in Israel (ABSL: Aronoff et al., 2005; Padden et al., 2010; Sandler et al., 2005), Kata Kolok in Indonesia (de Vos, 2011, 2014), Adamorobe Sign Language in Ghana (Nyst, 2010), and Martha's Vineyard Sign Language in the USA (Groce, 1985). Village sign languages differ consistently with community sign languages along three principle social factors: size of the community, insularity of the community, and the presence of hearing language users in the signing community. These factors can produce linguistic systems that evolve structures at very different rates.

For instance, ABSL is a village sign language that arose in the Negev desert region of Israel in a small, rural bedouin community. The signing community can be traced back to the founder of the village community and a high rate of hereditary deafness has persisted through consanguinous marriage in the insular community. As of 2010, there were 3,258 individuals in the bedouin community, approximately 150 of which were deaf signers (Meir, Assaf et al., 2005; Sandler et al., 2005). Over three generations of signers, ABSL has developed robust word order (Sandler et al., 2005) and productive affixation (Meir, Aronoff, Sandler & Padden, 2010), as well as the beginnings of a consistent noun-verb distinction (Tkachman & Sandler, 2013). However, studies of ABSL have also highlighted the fundamental linguistic structures that it lacks. ABSL does not exhibit duality of patterning, suggesting that it does not yet have a phonological system (Sandler et al., 2011), and lacks a consistent spatial agreement system (Padden et al., 2010). The evidence from ABSL, covering a wide range of linguistic domains, suggest that complex grammatical structures take time to emerge, and that the initial community alone is not enough to produce a fully linguistic grammar.

In particular, the findings from ABSL contrast with data from ISL, a language of a similar age which is found in a similar geographical region. ISL, in contrast to ABSL, does demonstrate a phonological system, and overall exhibits more regular linguistic constructions than does ABSL. A suggested reason for the difference between the linguistic structures found in these two languages is the differences in their social structures, exemplifying the distinction between community sign languages and village sign languages (Meir, Assaf et al., 2005). ISL has a larger, more dispersed and more diverse population. ABSL, on the other hand, emerged in a small community, based around the family unit; the formation of sub-systems in ABSL based around the family (so-called ”familylects”) attest to the importance of the family as a social unit (Sandler et al., 2011). The small, close-knit make up of the ABSL community has been suggested to affect the community structure, such that shared knowledge within families allows communication to be less explicit, or expressive, and thus may lead to greater linguistic irregularities as signers rely more on contextual information (Meir, Assaf et al., 2005; Meir, Sandler et al., 2005). The differences between village sign languages and community sign languages point to the impact of the community on linguistic structure; primarily, it highlights the crucial roles that use of a language in the community and transmission of that language over generations have in shaping language.

The study of emerging sign languages allows the observation of the structural features that emerge early on in a linguistic system, and offers the opportunity to track linguistic structures in a new system as they develop through use in a community, and as they are learnt by new members of that community. Furthermore, the typological differences between sign languages that reflect differences in the socio-cultural context bring to light the vast impact that cultural mechanisms have on linguistic structure. As the only current evidence of truly novel linguistic systems that occur naturally, emerging sign languages offer a unique window on the early evolution of a linguistic system.

1.4.4 The evolutionary stages of language

The body of sign language research discussed here displays the impact that such work has had on our understanding of language emergence, as well as of language more generally. Studies of manual communication systems highlight those features of language that are most common cross-linguistically and cross-modally, and further point to the features of languages that are governed by physical modality.

For the field of evolutionary linguistics, sign language research has been essential. Sign language studies allow the observation of languages that are, on the whole, younger than spoken languages under linguistic investigation. Furthermore, the study of restricted manual systems, such as homesign and emerging sign languages, offer a view of language in its earliest stages and provide a window onto the evolutionary trajectory of language emergence. Homesign systems evidence the capacity for linguistic communication in the individual, and indicate those 'resilient' properties of language that demonstrate underlying cognitive preferences in the individual. That these systems do not evolve beyond the restricted context within which they develop suggests that individuals do not make language: a community is required to make the leap to linguistic structure. Emerging sign languages provide a view of this leap from non-linguistic communication to linguistic system. The formation of a new community allows the emergence, and, over time, evolution of a linguistic system. Observing the first crucial generations of a linguistic system provides behavioural evidence of how such a system adapts to the needs of its users. Finally, studying established sign languages in contrast with restricted systems demonstrates how a linguistic system regularises and systematises as its community changes, with the entry of new members into the population. Spoken language systems at each of these stages do not exist in natural environments, and so the study of manual communication systems is essential to understanding the natural emergence of linguistic structures in the individual, in the newly-formed community, and in the established linguistic community.

Investigations of manual communication systems have provided unique insights into the roles that interaction in a community and transmission to new members of the community play in the emergence of linguistic structure, as well as the role that particular social structures can have in shaping language. Sign language research provides data for these phenomena in a natural linguistic setting, contributing to the typological and theoretical linguistic landscape. However, the study of natural language systems limits our ability to tease apart and contrast the roles that individual mechanisms play. Robustly comparable natural language environments are rare, making it difficult to understand in-depth how a particular social mechanism or structure may shape language. An experimental framework informed by natural language data can shed light where natural language data cannot reach, by allowing precise manipulations of the communicative context (i.e. the interactive and transmissive mechanisms that are at play), or the linguistic feature of interest. Recent work in the manual modality, which I review in the following section, adds to previous experimental work in artificial language learning, and aims to complement natural language data, using data from emerging sign languages to probe the cognitive preferences that underlie typological patterns cross-linguistically.

1.5 Bridging the gap between laboratory and field

Silent gesture studies present an experimental paradigm that follows similar motivations to artificial language learning experiments, testing the cognitive preferences of the individual in the production of communicative signals, but using the manual modality to reduce interference from existing linguistic knowledge on participants' productions. Hearing participants, who have no knowledge of sign language, are asked to communicate a set of concepts or events using only gesture. Silent gesture exploits a modality which can be used linguistically (i.e. in sign languages), but which is not normally used linguistically by participants in these experiments. As such, it allows linguists to probe preferences for particular linguistic structures, reducing the effect that any previously learnt language might have on their responses. Gestures produced as primary communication by participants in silent gesture experiments (where the hands are the primary articulators) demonstrate significantly different structures than when hearing language users produce co-speech gesture (Singleton, Goldin-Meadow & McNeill, 1995), and the gestures participants produce do not exhibit the structural properties of their native language, but offer evidence of natural cognitive preferences (Özcalışkan, Lucero & Goldin-Meadow, 2016).

Like artificial language learning experiments, silent gesture experiments have, in the majority of their instantiations, attempted to investigate how cognitive biases in the individual give rise to the typological patterns we see cross-linguistically. However, silent gesture studies have also been heavily informed, and motivated, by sign language data, and have demonstrated an attempt to bridge the gap between the documentation of natural sign languages and experimental linguistic research. In particular, silent gesture research has tested the preferences of participants for event ordering, along two main investigative lines: the segmentation of event gestures, and basic signal ordering patterns.

1.5.1 Segmentation and simultaneity in silent gesture

Sign languages are manual-visual languages that are able to exploit iconicity to convey information about states, objects and events in the real world. As such, sign languages exhibit simultaneous structures to an extent not found in spoken language, in which the physiology of the human speech apparatus prevents simultaneous articulation of multiple speech units. For example, it is possible to manually represent a ball rolling down a hill by simultaneously signalling the manner of the action (rolling) and the path of the action (down a hill), whereas in English we must sequence the two characteristic aspects of the event one after the other. Simultaneity is considered a natural property of manual systems, and the conflation of manner and path in describing motion events has been attested in co-speech gestures, homesign systems and sign languages (A. Senghas et al., 2004; A. Senghas, Ozyürek & Goldin-Meadow, 2013).

Signers in the first cohort of NSL produced holistic signs for motion events, conflating manner and path in comparable proportions to hearing co-speech gesturers. However, from the second cohort, A. Senghas et al. (2004) recorded a drastic reduction in the use of simultaneous structures in preference for segmented, sequential structures, such that gestures depicting the manner of movement and the path of the movement were articulated separately. The authors suggest that this is evidence of a system demonstrating increasingly linguistic structure, exhibiting two properties considered to be 'design features' of language, discreteness and combinatoriality (Hockett, 1960).

Silent gesture studies following this work aimed to understand the extent to which this change in NSL reflected an underlying preference for segmented structure, or to what extent sequential structure emerged against natural, iconic representations. Investigations into the representation of motion events in homesigners and silent gesturers have shown that silent gesturers preferentially use simultaneous structures over sequential structures when gesturing (Clay et al., 2014; Goldin-Meadow, 2015; Özyürek, Furman & Goldin-Meadow, 2014; Smith, Abramova, Cartmill & Kirby, under revision). Homesigners, on the other hand, show some combination of simultaneous and sequential structures, producing conflated manner-path gestures with an additional manner or path gesture in sequence. Clay et al. (2014) also contrasted silent gesture production of motion events in adults and children, finding that children produced significantly fewer simultaneous structures than adult gesturers. Children, in this case, demonstrate a greater preference for sequencing and segmentation than do adults, though it must be noted that the child gesturers do still show the greatest preference for simultaneously structured gestures. An important insight from these studies is the lack of segmentation in silent gestures and homesigners, in comparison to segmentation in full languages. Though homesigners show increasing segmentation of gesture sequences for motion events, the continued use of manual communication as the primary communication system in an individual is not sufficient for discrete, combinatorial structures to emerge. The silent gesture research, alongside sign language data, points to an evolution, over time, from simultaneous to sequential structure. Indeed, older sign languages such as ASL do make use of more sequential structures than emerging sign languages (Aronoff et al., 2005). Smith et al. (under revision) investigated this evolutionary development, modelling the transmission of signals for motion events over generations of an artificial sign language in an iterated learning framework. Hearing participants were first trained on gestures produced by a previous participant and then asked to produce their own gestures to describe a set of motion events. Through the process of transmission, Smith et al. (under revision) showed that weak preferences for sequential structures could be amplified to become the dominant strategy for representing motion events within a chain (see section 1.5.3 for further discussion of silent gesture in a cultural evolutionary framework).

The use of silent gesture experiments to further understand the emergence of sequential structure in language presents a methodology informed by data from natural sign languages, but one which produces behavioural evidence for the preferences that individuals have when innovating communicative signals, and how these preferences differ from those produced within a linguistic community.

1.5.2 Silent gesture and the natural order of events

Another principal area of research in silent gesture studies is examining preferences for word orders, and how such preferences map onto typological patterns cross-linguistically. Of the six possible word orders based on basic ordering of subject (S), object (O) and verb (V), two word orders, SOV and SVO, are overwhelmingly prevalent as basic word orders cross-linguistically (Dryer, 2011). Sign languages demonstrate similar preferences; a survey of 42 sign languages (established sign languages as well as emerging sign languages) found that SOV and SVO are preferred cross-linguistically, and are universally grammatical across the sampled languages (Napoli & Sutton-Spence, 2014). A wide range of silent gesture experiments have attempted to understand the motivating factors that can explain particular word order patterns. In an early instantiation of this work, Gershkoff-Stowe and Goldin-Meadow (2002) found that English-speaking silent gesturers preferred verb-final ordering patterns. Goldin-Meadow, So, Ozyürek and Mylander (2008) replicated this result in English-speaking participants, as well as showing a similar preference among speakers of very different languages (Turkish, Spanish and Chinese), where participants across all languages not only preferred verb-final structures but showed a dominant preference for SOV order.

Silent gesture studies following these foundational experiments have sought to further understand how the structures of events in the world may map onto linguistic structure. Schouwstra and de Swart (2014) analysed gesture productions for events that differed in terms of semantic constructions: whether the events were extensional (involving action through space, such as *carry*), or intensional (involving an object that is either non-specific, non-existent or dependent on the action, e.g. *want*, *dream*, or *build*). Participants presented with events of each kind demonstrated differently ordered sequences depending on the event type. Extensional events, like those used by Goldin-Meadow et al. (2008) and Gershkoff-Stowe and Goldin-Meadow (2002), produced gesture sequences of SOV order. Intensional events, however, were overwhelmingly produced with SVO ordering, demonstrating the semantic conditioning of word order. Importantly, this conditioning occurred independently of native language word order, in speakers of Dutch (an SVO language) and Turkish (an SOV language).

Other investigations into the origins of word order patterns in natural languages have suggested, based on results from silent gesture experiments, various explanations for these word order preferences. Hall, Mayberry and Ferreira (2013) assert that constraints against role conflict, based on the gesturer using their body to iconically represent agents of sentences, lead to separation of subject and object when two animate participants are involved in the event, and thus produce more SVO structures. Their findings, which evidence SOV in events with inanimate patients and SVO in events where both subject and object are animate, are supported by Meir et al. (2014), who find a similar pattern in silent gesturers and signers from three emerging sign languages: ABSL, ISL, Kafr Qasem Sign Language. However, Meir et al. (2014) offer a different explanation, suggesting that an overall preference for placing animate arguments first in a sequence is indicative of a general animacy bias that can account for the same patterns found by Hall et al. (2013). Many of these motivational arguments propose similar factors, principally that the need to disambiguate grammatical roles of animate agents and patients motivates the presence of SVO orders (Futrell et al., 2014; Gibson et al., 2013; Langus & Nespor, 2010; Marno et al., 2015). These different accounts for word order preferences across different experimental contexts underscore the insight offered by silent gesture studies in understanding how cognitive preferences in the individual are reflected in linguistic representations. In the absence of an existing set of conventions, utterances are structured flexibly, on the basis of the semantic properties of the event being described.

Most of these accounts assert that the gestures produced by silent gesturers represent natural preferences for event ordering. More recent work, contrasting individual gesturers with participants under cultural pressures for interaction (Christensen, Fusaroli & Tylén, 2016; Schouwstra et al., 2016), has assessed how natural preferences change when subjected to functional pressures for communication. Christensen et al. (2016) replicate the semantic conditioning evidenced by Schouwstra and de Swart (2014), but further contrast production in individuals with production in communicating pairs. Their results suggest that participants align gesture orders with their partner, potentially shifting 'natural' word-order preferences during communication. Finally, they manipulated the frequency of the types of events participants saw, and found that conventionalisation through communication led to dominance of the more frequent order. Schouwstra et al. (2016) also found that communicating participants regularise word order through conventionalisation, a trend that also presented itself in a second manipulation where pairs of participants communicated with each other, but where one participant was replaced every round, modelling population turnover of language learners. This work, as well as the experiments conducted by Christensen et al. (2016), have shown that natural preferences for event ordering in individuals (in the absence of a conventional communicative system) are not always fully reflected in the structures that emerge during communication and transmission. Where improvising individuals prefer to condition their word order usage on semantic properties of an event, this tendency disappears over time (under the influence of interaction as well as transmission) in favour of more regular word order.

The case of uncovering preferences for word order patterns has provided fertile ground for silent gestures studies, allowing the investigation into the precise pressures that lead individuals to produce particular patterns, and how these patterns map onto typological data from natural languages. Furthermore, this body of work has demonstrated the incorporation of silent gesture experiments into a cultural evolutionary framework, presenting work that attempts to understand how communication and learning affect individual preferences for word order patterns.

1.5.3 The cultural evolution of silent gesture

The two major lines of silent gesture research discussed above suggest how experimental work in the manual modality, informed by natural language data, can shed light on the cognitive mechanisms that drive the emergence of linguistic structures. Silent gesture paradigms have been implemented to investigate a variety of linguistic phenomena in addition to those discussed; for example, spatial reference (So, Coppola, Licciardello & Goldin-Meadow, 2005), temporal constructions (Schouwstra, 2016) and the systematic patterning of iconic strategies to represent the animacy and agency features of a referent (Hwang et al., 2016; Padden et al., 2013). For the most part, silent gesture studies have focussed on production in the individual, implementing tasks in which individual participants must innovate gestural signals de novo. However, more recently there have been moves to place silent gesture into a cultural evolutionary framework, to understand the impact of interaction and transmission on the gestures participants produce.

As discussed above, Christensen et al. (2016) found that pressures for interaction led to greater conventionalisation of word order strategies, reducing the effect of semantic, or iconic, conditioning. Other research that has analysed silent gesture in an interactive context has demonstrated the efficacy of iconic gestures in bootstrapping meaning in an innovated communication system (Fay et al., 2013), and has evinced the role that conventionalisation in interaction plays in changing formal properties of innovated gestures (Namboodiripad, Lenzen, Lepic & Verhoef, 2016). Namboodiripad et al. (2016) presented a novel method that analysed silent gestures in interaction using a Microsoft Kinect device to record information about joint-use whilst gesturing. Their results exhibited a reduction in the size of gestures, and in the length of gesture sequences, over rounds of interaction, underscoring Clark's "principle of least collaborative effort" (1986), such that participants negotiate a system that reduces production effort whilst retaining communicative accuracy. Rarer are studies that attempt to model the transmission of silent gesture systems to new learners. Smith et al. (under revision) modelled the effect of transmission to new learners in a silent gesture iterated learning experiment that analysed gestures produced to convey motion events, finding that the transmission process could amplify a weak preference for segmented, sequential structures in the gestures their participants produced. Schouwstra et al. (2016) contrasted pairs of participants who either communicated with each other over several rounds (interaction-only), or interacted with gradual replacement of a participant at every round (combining interaction and transmission) and found that both conditions led to conventional ordering patterns of gesture sequences, suggesting that interaction alone is sufficient for the regularisation of word order patterns.

This nascent area of silent gesture research has clearly highlighted the importance of accounting for interaction and transmission in the development of manual communication systems. However, cultural evolutionary models of manual communication are still in their infancy. Studying gesture presents a number of methodological problems that are exacerbated when interaction and transmission are incorporated into the model, such as recording data and measuring features of gestures. However, for silent gesture to really bridge the gap between experimental work and natural language data, modelling natural linguistic processes is essential. In particular, data from emerging sign languages offers a window onto the early stages of linguistic evolution, but would benefit from support from a constrained and manipulable experimental framework that allows the study of particular linguistic phenomena under particular evolutionary pressures. This thesis presents a novel method that attempts to clarify the research carried out on naturally emerging systems, by utilising precise manipulations of linguistic features and cultural contexts.

1.6 Artificial sign language learning

1.6.1 A novel method for evolutionary linguistics

The experimental work I present in the remainder of this thesis aims to investigate the precise roles that the cultural evolutionary mechanisms of interaction and transmission play in the emergence and evolution of linguistic structure, seeking to understand how communities of learners shape language. Across the set of experiments presented here, individual learners without a conventionalised input model innovate gestures to communicate concepts and events. Pairs of participants communicate using gesture in an interactive setting, and the gestures participants produce are transmitted to new learners, modelling the generational transmission of language. I present a set of experiments motivated and informed by natural language data, particularly recent work into emerging sign languages, to model the cultural evolution of language in the manual modality, from innovated pantomime to systematic, language-like signalling systems.

The methodology I describe throughout this work draws from a combination of artificial language learning, silent gesture studies, and experimental work using referential communication tasks and iterated learning methods. Participants produce gestures, use gesture to communicate, and learn gestures produced by other participants. Participants learn artificial sign languages used to convey a set of meanings, and the analyses presented here measure how these systems change through the use of the signals in interaction and through the transmission of the signals to new learners. This work builds substantially on previous artificial language learning research, discussed in section 1.3.2, and the silent gesture work described in section 1.5, allowing the study of artificial sign languages in a manipulable and constrained environment that can be used to test the emergence of particular linguistic structures. These experiments are designed to complement emerging sign language research on the evolution of linguistic structure, which is limited by the natural, complex environment which otherwise makes it so valuable to linguistic study.

Building a cultural evolutionary framework for the manual modality further offers insight into the evolution of linguistic structure cross-modally. By comparing my experimental results with findings from previous interaction and iterated learning experiments, my work elucidates those processes and properties of language that are common across modalities. However, this paradigm can be applied (and is here) to understand modality-specific forms of communication, to further understand how physical and cognitive demands of the modality lead to different linguistic signalling forms.

Indeed, I address linguistic phenomena that are considered universal properties of language, such as grammatical categories, which may manifest themselves differently in the manual modality and in speech. I also investigate the emergence of a phenomenon considered to be unique to sign languages, spatial agreement, to understand how modality-general cultural evolutionary mechanisms impact a feature conditioned by the possibilities of visual representation.

The artificial sign language learning paradigm I present in this work offers a novel method for precisely understanding the impact of particular cultural evolutionary mechanisms on the evolution of linguistic structure in the early stages of language. It provides a substantial expansion on previous research in the cultural evolution of manual communication systems, and lights the way for further research into how linguistic systems in the manual modality change over time, to answer fundamental questions about the human capacity for producing and learning complex, structured language.

1.6.2 Thesis outline

The majority of this thesis is comprised of a set of artificial sign language learning experiments that demonstrate the method and provide empirical insight into the cultural evolution of language in the manual modality.

Chapter 2 describes an experiment that questions the precise roles that interaction and transmission play in the evolution of systematicity in language. I contrast participants in a transmission plus interaction condition with an interaction-only condition and a transmission-only condition, to tease apart the role of each mechanism. The results of this study suggest that neither interaction alone nor transmission alone is sufficient for the emergence of systematic, language-like signals. Rather, both mechanisms in combination lead to systems of signals that are both systematic and communicatively efficient.

In chapter 3, I further investigate the impact of interaction in the emergence of grammatical categories, specifically in the emergence of noun and verb distinctions. The experiment detailed here is more directly motivated by data from emerging sign languages than that described in chapter 2, using stimuli employed previously to elicit linguistic constructions from signers of NSL, as well as Nicaraguan homesigners (Abner, Flaherty, Stangl, Brentari & Goldin-Meadow, 2015; Abner et al., 2016), thus allowing comparability with natural language data. Though the results from this study do show evidence of the beginnings of systematic noun and verb distinctions, they also indicate, in accordance with the results from chapter 2, that interaction alone is not sufficient for the emergence of systematic category structures, though the gestures participants produce become conventionalised within a pair.

Chapter 4 provides an experimental account of the evolution of spatial verb agreement, following research into emerging sign languages that suggests that complex spatial constructions take time to emerge (Meir et al., 2007; Padden et al., 2010). This experiment also expands on previous iterated learning experiments to address the emergence of a complex grammatical construction. Furthermore, focussing on a feature strongly considered to be governed by the manual modality, I assess how the ability to represent grammatical relations visually affects the emergence of systematicity. Participants' gestures exhibit systematic strategies that distinguish sentence participants from one another, and in several cases, do this using the space around the gesturer. The evolution of these gestures over generations in a transmission chain and through interaction within a pair shows similar trajectories to the evolution of spatial grammar in two young sign languages, ABSL and ISL (Meir et al., 2007; Padden et al., 2010), but still retain high levels of iconicity. These findings also shed light on a larger debate in sign language linguistics, about the integration of gestural elements into a linguistic system.

Finally, chapter 5 brings together the threads of each experimental setup to discuss the contribution of this work to our understanding of the cultural evolution of language, and how experimental investigations in the manual modality contribute to linguistic enquiry more generally. I discuss the importance of the methodology I present throughout the thesis, and suggest future applications that could further elucidate the effect that different socio-cultural factors may have on the emergence and evolution of systematic, linguistic structure.

Chapter 2

Evolving artificial sign languages in the lab: from improvised gesture to systematic sign

2.1 Motamedi et al. (under revision): author contributions

Material in this chapter has been reproduced from Motamedi et al. (under revision), submitted to the journal *Cognition*[1]. The paper was co-authored with Simon Kirby, Marieke Schouwstra, Kenny Smith and Jennifer Culbertson, and has been reproduced here with the permission of all authors. The experiments described here were discussed during supervision meetings, and all authors contributed to the writing and editing of the paper. Analyses from section 2.6 onwards have been adapted from the supplementary materials submitted with Motamedi et al. (under revision).

2.2 Introduction

Languages exhibit systematicity; single utterances are not isolated, independent units but form part of a structured system of interdependent elements. We see systematicity across levels of language, in the lexicon, morphology and syntax. Parts of signals are re-used and recombined across utterances and correspond systematically to different aspects of the meanings being conveyed. For example, the noun phrases *blue shoes* and *red shoes* both include *shoes*, indicating which part of their meaning

[1] This paper is now under revision, following peer review that is part of the submission process for the journal *Cognition*. The proposed revisions are substantial and include having the gestural data blind coded by a second coder, and running further variations of the transmission plus interaction and interaction only conditions without a pressure for fast communication.

is shared, and differ on their descriptors, *blue* and *red*, this difference in form signalling a difference in meaning. One of the central challenges of language evolution research is to explain how systematic structure arises in language, and which mechanisms drive its evolution. The prevalence of systematic structure across languages and modalities points to its status as a fundamental property of language. Cultural evolutionary accounts propose that systematic structure develops as language adapts to pressures arising from language use and the transmission of language to new learners (Cornish et al., 2013; Kirby et al., 2008, 2015; Silvey et al., 2014). However, investigation of such cultural processes requires the observation of communication systems at different stages of linguistic emergence, in order to understand how these mechanisms affect languages at different evolutionary stages. Currently, there are two main sources of evidence available for such observations: emerging sign systems, providing data from natural languages which are in the early stages of developing linguistic structure, and experimental research modelling language early in its evolutionary development. We draw from both of these sources to inform our approach, observing the emergence of manual communication systems in the laboratory with a focus on the effects that the cultural mechanisms of interaction and transmission have on the the evolution of these systems. In essence, we create a controlled environment in which we can observe the evolution of minature artificial sign languages.

2.2.1 Field research: Homesign and emerging languages

Observations from homesign (Goldin-Meadow, 2003; Haviland, 2013) and emerging sign languages (Aronoff et al., 2005; de Vos, 2014; A. Senghas et al., 2004) provide the only naturally occurring evidence of language at its earliest stages, and provide crucial insights into the different cultural contexts that affect the structure found in communicative systems. Homesign systems show children creating communication systems when they are deprived of linguistic input. These manual communication systems are created by deaf children (usually born to hearing parents) who are not exposed to an accessible conventional language early in their development, and must improvise ways to communicate. Homesigns exhibit some structural properties in common with established languages, such as regularities in syntax (Goldin-Meadow & Feldman, 1977), morphology (Goldin-Meadow et al., 1995) and lexical categorisation (Goldin-Meadow et al., 1994; Haviland, 2013), but do not reflect parental input, either from spoken language or infant-directed gestures (Goldin-Meadow & Mylander, 1983, 1998). Despite these structural properties, homesigns differ from established languages in a number of ways. They show less systematic structure and less regularity than sign languages (Goldin-Meadow et al., 2014), and lack consistency across users of a single system (Richie et al., 2014). Critically, homesign systems are developed by individuals and used within the family of the deaf individual. Hearing family members may also use the system, but use of homesigns in communication is often limited and asymmetrical, with only the deaf family members using homesigns as their primary communication system. Homesign systems typically persist only for a single generation; lack of a community halts further transmission, and thus evolution, of the system.

Observation of early sign languages as they develop has also illustrated the effects that interaction within a community and transmission to new community members have on the form of the language. Emerging sign languages arise when communities are formed by deaf individuals, who lack a conventional language model, or who are otherwise cut off from pre-existing languages. In many cases of sign language emergence, linguistic systems begin as improvised homesigns within family units and develop as they are learnt by, and transmitted to, a wider community. These communities may emerge due to high rates of hereditary deafness, as is the case with many village sign languages such as Al-Sayyid Bedouin Sign Language in Israel (Aronoff et al., 2005), Kata Kolok in Indonesia (de Vos, 2014) and Adamarobe Sign Language in Ghana (Nyst, 2010), or they may emerge due to changes in educational policy, such as the provision of schools for the deaf. The latter social context provided ground for the development of Nicaraguan Sign Language (NSL), which emerged in the late 1970s after a deaf school was established in Managua, and deaf individuals who had developed their own homesign systems were then able to interact with each other and develop a conventionalised language across the school community (Kegl, Senghas & Coppola, 1999; A. Senghas & Coppola, 2001; A. Senghas et al., 2004). Studies of emerging sign languages have shown how early sign systems might develop linguistic features such as conventional word order (Sandler et al., 2005) and role shift (where signers take on the role of another 'character' in the discourse) (Kocab, Pyers & Senghas, 2014), and how such structures change over time. A. Senghas et al. (2004) present data from NSL that demonstrates how motion events that were signed holistically (conflating manner and path) in the first cohort of signers became sequential (separating manner and path signals) in later cohorts. Their results are particularly surprising, as simultaneous structure is a common modality-specific property of sign language morphology, which allows iconic event representation (Aronoff et al., 2005; A. Senghas et al., 2004); as such, it is striking that new learners of NSL do not exploit this iconicity.

Goldin-Meadow et al. (2014) similarly examine the evolution of NSL, analysing the consistency of handshape for nominals and predicates in four groups of signers:

Nicaraguan homesigners, signers of NSL cohort 1, signers of NSL cohort 2, and ASL signers. Their results suggest that ASL, the oldest and most stable language in the sample, exhibits the most consistent handshapes for nominals and predicates across signers. Homesigners, at the other end of the scale, exhibited low consistency in the handshapes used across the group, as would be expected from individual innovators, whilst the two cohorts of NSL signers show a progressive increase in the use of consistent handshapes across signers. These results point to the importance of a community in the development of a language. Use of a system within the community allows signs to conventionalise and become more regular, and transmission to new members of a community entrenches this process further; repeated use and transmission of a system leads to increasing consistency across signers.

The early sign systems described above provide a valuable perspective on language emergence, allowing observation of how different cultural contexts (for example, access to a community and the growth of a community through the introduction of new learners) affect the structure present in linguistic systems (A. Senghas, 2005). Although crucial to our understanding of language evolution, natural language observation requires large-scale longitudinal study of phenomena generally outside the researcher's control. Experimental research, on the other hand, presents the opportunity to manipulate both the linguistic structures and the social environments we wish to investigate, and constitutes a valuable companion to field research.

2.2.2 Experimental research: communication, iteration and gesture

Experimental semiotic studies using graphical communication tasks have focused on how interaction shapes communication. Participants in these studies took part in a communication game similar to the board game *Pictionary*, in which they had to communicate concepts by drawing them whilst other participants attempted to interpret the drawings (Fay et al., 2010; Garrod et al., 2007; Healey et al., 2007). Findings from this paradigm have demonstrated that repeated interaction leads to an increase in the production of symbolic signals. The drawings participants produced become smaller, less iconic and less complex, leading to more efficient and more successful communication.

The iterated learning paradigm tests the effect of transmission of a system to new learners on the structure of a linguistic system. In an iterated learning experiment, initial participants are trained on an artificial language and are subsequently asked to reproduce the language they have learnt. Their reproductions are then passed on as the input language for the next participant, who then attempts to reproduce what they have learnt. This process is repeated, with the output from one participant used as input for the next, modelling a process analogous to the generational transmission of language. Experiments using the iterated learning model have shown that systematic structure emerges through this process, suggesting that such structure develops in response to pressure for communication systems to be learnable (Carr et al., 2016; Kirby et al., 2008, 2014, 2015).

More recently, a body of work has probed the particular roles that interaction and transmission play together. Kirby et al. (2015) and Carr et al. (2016) compare systems built by interaction alone with those built by a combination of interaction and transmission. Participants either took part in closed pairs, in which they communicated repeatedly without the introduction of new learners, or in chains of pairs, where one pair learned from the previous pair. Signals produced by chains of pairs evolved to be both learnable and useable in communication, as opposed to signals produced by closed pairs, which became efficient for communication, but lacked learnable structure. Theisen-White et al. (2011), amongst others (Caldwell & Smith, 2012; Fay & Ellison, 2013) demonstrate similar results in the graphical medium, combining iterated learning and a graphical communication task to investigate how both interaction between participants and transmission to naive participants may lead to the emergence of structured signals. Pairs of participants played a graphical communication game, communicating about a set of concepts that are associated on two dimensions: a thematic dimension and an entity type dimension. Concepts in the meaning space shared properties on either the thematic dimension (such as 'doctor' and 'hospital') or the entity type dimension (such as 'hospital' and 'school'). Pairs of participants introduce some systematic structure in their signals, reflecting the structure found in the concepts they had to communicate. For example, for the meanings of 'doctor' and 'hospital', participants re-used drawings that showed some aspect of medicine, such as a stethoscope. However, it was transmission of the signals to new pairs that led to a cumulative increase in the systematic structure present in the drawings participants produced, over the course of five pairs, or 'generations', in the experiment. Results across these studies suggest that, although individual mechanisms may shape language, it is the combination of pressures from interaction and transmission that leads to structures that are both efficient and systematic, suited both to communication and learning.

Experiments using silent gesture investigate the cognitive constraints that might affect the structures produced in language. Silent gesture studies use hearing participants with no knowledge of any sign language and ask them to communicate using gesture but no speech. The gestural modality reduces the confounds that existing linguistic knowledge may introduce, by using a non-linguistic but communicative medium, and provides a useful tool for probing preferences for different linguistic properties. For example, Goldin-Meadow et al. (2008), amongst others (Hall et al., 2013; Schouwstra & de Swart, 2014) used gestures from hearing participants to probe the cognitive biases that influence event ordering, showing that participants from different linguistic backgrounds overwhelmingly produce the same word order when gesturing, favouring verb-final structures. Smith et al. (under revision) combine silent gesture with iterated learning to investigate how the expression of motion events develops through the transmission of emerging sign systems. This study offers an experimental comparison with the findings concerning motion events in NSL, where signers from the second cohort begin to sign motion events using segmented structure, though previous cohorts had signed motion events simultaneously (A. Senghas et al., 2004). Participants produced gestures about a set of videos showing a ball moving in varying manners (e.g. bouncing or rolling) and along different paths (e.g. a slope or a circular path). The gestures individual participants produced were recorded and used to train further participants, who were then asked to produce gestures for the same set of scenes. As the gestural systems were transmitted, they became increasingly systematic, showing increasing regularity in the forms of the gestures used. Smith et al. (under revision) found that, although simultaneous structures were favoured, some sequential structures did emerge. In these cases, sequential structures could be amplified by the transmission process, becoming frequent in a particular chain; or, they could be ignored in favour of simultaneous structures that, through transmission, became more and more frequent. Silent gesture can offer insight into the cognitive biases that affect the structures we produce as individuals; combining such a tool with iterated learning can further inform us about how preferences in the individual may be amplified and become persistent in a community.

2.2.3 From pantomime to sign language

We aim to investigate how language emerges from non-linguistic communication and understand the roles that transmission and interaction play in its development. We model this evolution from pantomime to sign language, focusing on how systematic structure emerges in manual systems, using silent gesture to understand the effects of interaction and transmission. We present a novel study that investigates the roles these mechanisms play in combination with each other, and what effects each mechanism has when isolated. At the earliest stages in the evolution of manual communication, signals are used unsystematically and independently of each other, akin to pantomime. For instance, the form of signals referencing similar meanings will not necessarily use similar properties, but might depict different salient aspects of the referent. Pantomimes exhibit high iconicity, taking up more space and requiring more effort (Klima & Bellugi, 1979). Additionally, they are not part of a conventionalised code, productions differing between communicators as well as between productions by the same individual (Klima & Bellugi, 1979; McNeill, 2000). Sign languages, in contrast, are coded, conventionalised systems used by a community. Linguistic signs are more efficient in form than non-linguistic signs (for example, by showing fewer redundancies, or being smaller), and signals are segmented and used systematically. Signs within linguistic systems are not independent of one another, but are re-used and re-combined consistently across the system to predictably differentiate between referents. Finally, although sign languages exploit iconicity to a degree, linguistic signs tend to be less iconic than pantomime, and are not transparent to non-users (Klima & Bellugi, 1979).

The present set of experiments investigates how interaction and transmission effect the emergence of language, from initial unstructured systems. We use silent gesture to model the emergence of manual communication, minimising interference from participants' existing linguistic knowledge. Use of the manual modality carries the benefit of comparability with naturally emerging languages, as our main source of emerging systems in natural environments are manual systems such as homesign and new sign languages (see section 2.2.1). Hearing participants without knowledge of sign language were asked to communicate about a set of concepts using only gesture. Experiment 1 investigates the effects of both interaction and transmission on the emergence of a manual communication system. Pairs of participants learned gestures produced by a previous participant and then communicated with each other using only gesture. The gestures they produced were used as training for a new pair of participants, who then had to communicate with each other about the same concepts, a process which repeated for five pairs, or 'generations'. Following Theisen et al. (2010) and Theisen-White et al. (2011), participants were asked to communicate about a structured set of concepts that could differ on two dimensions.

The present design allows us to investigate how systematic structure emerges over time in manual communication systems by posing the following questions: what do spontaneous gestures look like when there is no model to learn from? What effect does communication within pairs have on the gestures participants produce? And how does transmission to naive learners affect the gestures produced? In experiment 1, we predict that initial participants producing gestures without a model will produce pantomime-like gestures, lacking systematicity. We predict that the combination of interaction and transmission will lead to systematic signals structured according to the dimensions of the meaning space, composed of sub-elements which signal the two dimensions along which the concepts are organised, and which are re-used across categories within these dimensions. The combined pressures of interaction and transmission will lead to sets of signals that are both more efficient for communication and more learnable (Kirby et al., 2015), driving the evolution of language-like, systematic gestures from pantomime. We predict that these sets of gestures will follow the linguistic pathways demonstrated in the naturally emerging languages, where repeated use and transmission of a system leads to the emergence of more regular and more systematic languages (Goldin-Meadow et al., 2014; Sandler et al., 2005; A. Senghas & Coppola, 2001).

2.3 Experiment 1: interaction and transmission

2.3.1 Methods and Materials

Participants took part in one of two stages of the experiment. Our first set of participants were recruited as seed participants (described in section 2.3.1), who recorded a single gesture for a single concept in the meaning space. Following collection of these seed gestures, participants in the main stage of the experiment took part in pairs, that were organised into transmission chains of five generations. The gestures recorded in the seed collection stage were used as initial training gestures for the transmission chains in the main experiment.

Participants

48 seed participants (17 male, 31 female, aged 18 to 33, median age 22) were recruited to record an initial set of videos, each participant recording a single video. Seed participants were unpaid volunteers recruited from the undergraduate and postgraduate population at the University of Edinburgh. A further 50 participants (14 male, 36 female, age 18 to 32, median age 20) were recruited for the main stage of the experiment, from the University of Edinburgh careers website. All participants were right-handed native English speakers, with no knowledge of sign language.

Materials

Participants were presented with items from a meaning space containing a total of 24 meanings (see figure 2.1). The items were selected to create two meaning dimensions: thematic and functional, following Theisen-White et al. (2011). Items in the meaning space can share an association on the thematic dimension (for example, *chef* and *restaurant* share the thematic dimension of cooking), or on the functional dimension (for example, *chef* and *photographer* share the functional dimension of person). Meanings were presented as orthographic words, in order to avoid the ambiguity of meaning that images or video stimuli may cause.

		Functional dimension			
		person	**location**	**object**	**action**
Thematic dimension	**food**	chef	restaurant	frying pan	to cook
	religion	vicar	church	bible	to preach
	photography	photographer	darkroom	camera	to take a photo
	music	singer	concert hall	microphone	to sing
	hair styling	hairdresser	hair salon	scissors	to give a haircut
	law enforcement	police officer	prison	handcuffs	to make an arrest

Figure 2.1: The meaning space. Concepts shown in the rows of figure 2.1 share the thematic association of particular professions, while items in columns share the functional associations of *person*, *location*, *object* and *action*.

For all stages of the study, participants were placed in individual experiment booths, in front of an Apple Thunderbolt monitor with an affixed Logitech webcam. Both monitor and webcam were connected to an Apple Macbook Air laptop running Psychopy (Peirce, 2007) and VideoBox (Kirby, 2016), custom software we developed to record and stream video between networked computers.

Procedure: seed collection

Our first group of participants, the seed participants, were asked to produce a gesture for a single concept from the meaning space. The aim of the seed collection was to provide a set of variable gestures that would act as the first training set for the transmission chains. Participants were told that they would be presented with a concept, and that they should communicate that concept using only gesture. Participants were instructed that they should not speak whilst doing the task, that they should not attempt to manually spell concepts, and that they should remain seated throughout the task. No audio was recorded at any stage of the experiment. All participants were given a practice trial to familiarise them with the recording process. In the practice trial, participants were presented with the concept *angry*, and asked to communicate this concept using only gesture. The target concept was shown onscreen and participants could begin to record their gesture by pressing the space bar. After pressing the space bar, participants were given a 3 second countdown on screen, to prepare them for the start of recording. During recording, participants were shown a live mirrored stream of themselves on the display, allowing them to monitor their gestures and make sure that they were in frame. Recording in the seed collection task was timed for 7 seconds and stopped automatically. Following the practice trial, participants were shown one item from the meaning space as a single, orthographic word on screen, and followed the same procedure as the practice trial. Seed participants did not see any of the other concepts from the meaning space. For each item in the meaning space, two seed videos were recorded, giving a total of 48 seed videos.

Procedure: main experiment

The 50 participants recruited for the main stage of the study were organised in pairs, into five transmission chains. Pairs of participants were taken through a two-part procedure in which they were trained on a set of gestures, and then asked to communicate with their partner, using only gesture. Participants were seated in individual experiment booths, and communication was enabled through video streaming between two networked computers. As in the seed collection stage, participants were instructed not to use spoken language or manual spelling, and to remain seated throughout the task. Figure 2.2 illustrates the structure of a tranmission chain in the experiment.

Participants within a pair were trained simultaneously, in separate experiment booths. During training, participants were presented with videos of another person gesturing and asked to identify the correct meaning of the gesture. Each training trial consisted of the presentation of a gesture video, with simultaneous presentation of the meaning grid (shown in figure 2.3). Participants could watch the video in full before making their guess, or interrupt play by making a guess. Participants made their guess by clicking on a word from the grid. Once a meaning had been selected, participants were given feedback: they were told whether they were correct or incorrect, and shown the correct interpretation of the gesture. The gesture video was then played again in full, without the opportunity to interrupt. Participants were subsequently asked to copy the gesture, and given a 3 second countdown to prepare themselves for recording. During recording, they were shown the mirrored live stream of themselves, as in the seed collection stage, though the duration of recording was not pre-set. Participants controlled when recording finished by pressing the space bar to end recording when they had finished their gesture. Each round of

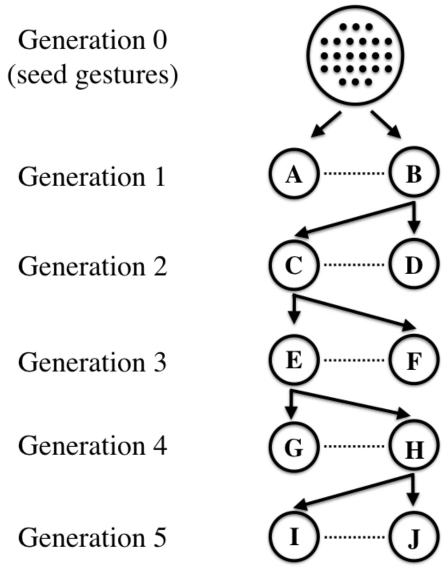


Figure 2.2: Transmission chain structure for experiment 1. Solid arrows represent transmission, dashed lines represent interaction. The seed gestures (generation 0) act as the starting sets for each chain. Participants in the first generation learn from the seed participants. They then communicate with each other and pass on their output (randomly selected from one of the two participants) as training for generation two. This process repeats for five generations in total.

training consisted of 18 trials, in which participants were trained on 18 out of the total 24 items in the meaning space.

The 18 training items were selected randomly at the start of training, and were balanced across both the thematic and functional dimensions, so that 3 items from each theme were used and either 4 or 5 items from each functional type. The same 18 items were presented in each round of training, and the same items were used as training for both participants in a pair throughout the training stage, though the order of presentation was randomised for each participant at each round of training. Participants completed 2 rounds of training, giving 36 trials in total. All 24 items in the meaning space were presented in the meaning grid used for interpretation of the gestures. For each participant, the position of items in the grid was randomised, but remained constant for the duration of the experiment.

For participants in generation 1, a training set was generated by selecting at random one of the two seed gestures for each meaning in the meaning space. In the results section, we refer to these sets of seed gestures as generation 0. For subsequent generations the training set consisted of gestures produced in the testing stage by the previous generation. Only gestures from one of the two participants were used as training, randomly selected at each generation, so that the full set of gestures from a

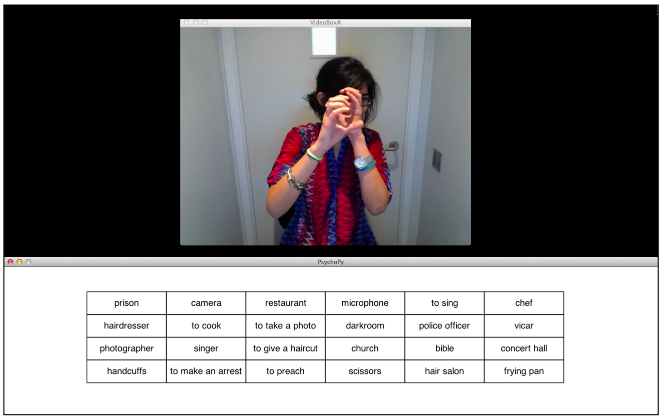


Figure 2.3: Screenshot example from a training trial, from a seed gesture for *camera*. The VideoBox window (top centre) presents pre-recorded videos in training as well as live stream during recording and testing. The psychopy window (bottom) presents the meaning grid for interpretation, and shows instructions and feedback.

single participant was transmitted and used as the model for the next generation. [2]

In the testing stage of the experiment, participants took turns to communicate (as director) and interpret (as matcher) items in the meaning space, with each participant in a pair producing a gesture for each meaning once and interpreting a gesture for each meaning once, for a total of 48 trials. The order of presentation for items in the meaning space was randomised.

As director, the participant was presented with an item from the meaning space. The director was given 3 seconds with only the target item on screen, followed by a 3 second countdown to prepare them for recording. The target meaning remained on screen throughout the trial. During recording, the participant performed their gesture, again seeing their own image mirrored onscreen, with a live, unmirrored stream sent to their partner's networked computer. The director was able to stop recording and turn off streaming by pressing the space bar at any time, at which point they had to wait for their partner's interpretation to continue. The matcher could also stop streaming at any time by pressing space bar (see below).

In the role of matcher, the participant had a short wait whilst the target item was presented to their partner, after which they were presented with a similar 3 second countdown to prepare them to receive the video stream from their partner. Once streaming began, the same grid of meanings that appeared in the training stage were shown on screen for the matcher to make their interpretation. However, they could

[2] We chose to transmit the output from one participant, instead of mixed output from both participants. Mixed training may obscure regularities in a participants' gestures, and could therefore complicate the learning process by presenting participants with conflicting training stimuli. Understanding learning from conflicting input investigates a different research question to that which we pose here (Singleton & Newport, 2004; Smith et al., 2016).

not make their guess (by clicking on an item in the grid) until streaming had stopped, either by the director stopping recording, or by the matcher themselves stopping streaming by pressing space bar. We included this step to ensure that timings recorded in the experiment would reflect the time took for a matcher to comprehend a gesture, and would not be confounded by participants spending time finding their responses in the grid of words. Once streaming had been terminated, the matcher was free to select their interpretation from the meaning grid. Following their interpretation, both the matcher and the director were given full feedback. They were shown whether the interpretation was correct or incorrect, and both the target item (which the director had been instructed to convey) and the meaning selected by the matcher were presented on screen orthographically, to both participants. Participants swapped roles after each trial, taking it in turns to be director and matcher for the duration of the study.

Throughout the testing stage, participants were shown a red timer on the right of the display, which ran when recording and streaming occurred and accumulated across all trials, in order to encourage fast and accurate communication. Participants were briefed about the timer and told that a cash prize would be offered to the pair that was both fastest to complete the experiment and which correctly interpreted the most gestures. The winner was judged by their overall time to complete the task with a three second penalty added for each incorrect interpretation, and participants were informed of how the winning pair would be judged in the experiment briefing. The timer was introduced to reduce participants' shyness and encourage naturalistic gestures, as a pilot ran in advance of the main experiment indicated that nervousness caused participants to produce gestures slowly.

2.3.2 Results: Qualitative data

Before detailing a quantitative analysis of the gesture sets, we present a qualitative analysis of the gestures produced over the different stages of the experiment.

Seed gestures

The gestures produced in the seed stage tend towards large, pantomime-like gestures and frequently depict elaborate scenes, such as that shown in figure 2.4, showing a gesture for the meaning *to make an arrest*. This example, typical of the seed gestures, exhibits the global structure found in pantomime.

Although many of these gestures are highly iconic, and often very elaborate, seed gestures sometimes lack features that could easily distinguish them from other meanings. For example, gestures produced for 'hairdresser','hair salon' and 'to give

(a) (b) (c)

Figure 2.4: Illustration of pantomime-like structure in seed gesture for *to make an arrest*. The participant illustrates a scene in which the presumed arresting officer draws and points a gun (a), before proceeding to run after the perpetrator (b), and then catching them (c).

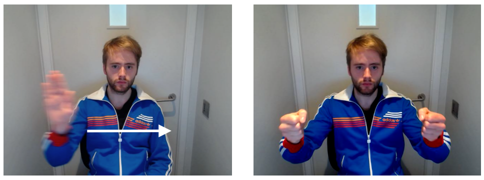
a haircut' (figure 2.5) all involve the gesturer miming cutting their own hair, but fail to provide any features that could allow an observer to distinguish between the three meanings. This is understandable considering that seed participants only see one meaning from the meaning space, and are not aware of the associated meanings that are part of the full set. These gestures, produced in isolation from each other, share a number of iconic features associated with their thematic category, but do not yet distinguish between meanings systematically.

(a) (b) (c)

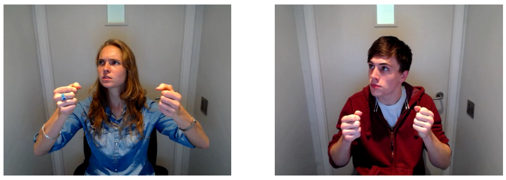
Figure 2.5: Ambiguity of gestures in the seed stage is exemplified by gestures for *hairdresser* (a), *hair salon* (b), and *to give a haircut* (c). All participants mime cutting their own hair, and there is little to distinguish between each meaning.

Transmission and interaction stage

The sets of gestures show change over the course of five generations, becoming more systematic and efficient. For example, the gestures in figure 2.6, taken from chain 3, show the productions for *prison* at generation 1 and gestures for the same meaning at generation 5. At generation 1, both participants gesture shaking the bars of a prison cell. These gestures, like their seed precursor, are holistic and iconic, a pantomime of being behind bars. However, unlike their seed model, the generation 1 gestures are shorter in length and involve fewer separate elements (both participants have dropped the palm movement seen in the seed gesture), exhibiting greater efficiency in their gestures.



(a) Generation 0 gesture for *prison*.



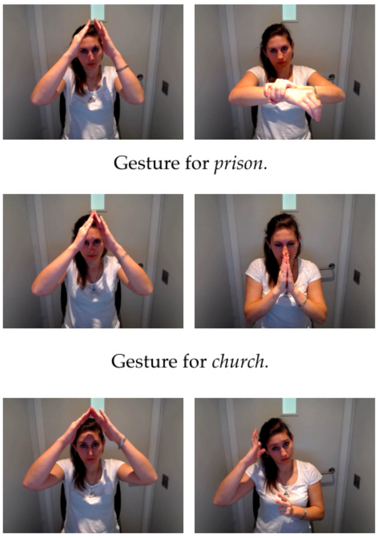
(b) Generation 1 gestures for *prison*. Gesture on left used as model for generation 2.

(c) Generation 5 gestures for *prison*.

Figure 2.6: Over generations (a, b, c) in a chain, gestures for *prison* become more systematic across participants. The participant at generation 0 (seed video) moves a palm from right to left before shaking the bars of prison cell. Participants in generation 1 repeat the bar-shaking gesture, but drop the palm movement in the seed gesture. However, by generation 5, the gestures have changed. Both participants use a roof gesture, followed by a cuffing gesture to communicate the same meaning.

In contrast, by generation 5 this meaning is gestured with a two-part sign. Both participants produce a roof gesture followed by a wrist-grabbing gesture indicating handcuffs. The structure of these gestures is no longer holistic, but segmented. Furthermore, segmentation appears to be widespread and systematic. For example, figure 2.7a gives examples of gestures from the same generation 5 participant shown in figure 2.6c, for the meanings *prison*, *church* and *hair salon*, all meanings from the *location* category. In each instance, the participant re-uses and recombines parts of signs; she consistently re-uses the roof sign as a category marker for location, followed by a thematic signal (in this case, a cuffing gesture for *prison*, a praying gesture for *church*, and a hair cutting gesture for *hair salon*. The set of gestures continues to develop from generation 1 onwards, indicating that changes in the system are not simply an effect of gestures being produced by the same person, but that this change is a continuous process.

Gestures also demonstrate the re-use of thematic signals across categories in the



Gesture for *prison.*

Gesture for *church.*

Gesture for *hair salon.*

(a) Re-use of the roof gesture across signals in the functional category of *location*.

Gesture for *photographer.*

Gesture for *camera.*

Gesture for *to take a photo.*

(b) Re-use of the camera gesture across signals in the thematic category of *photography*.

Figure 2.7: Segmentation into sequential, systematically re-used parts along the functional category (a), and the thematic category (b). Gestures in (a) communicate *prison*, *church*, and *hair salon*, and were produced by a participant in chain 3, generation 5, and show use of the same location marker, a roof shape. Gestures in (b) communicate *photographer*, *camera* and *to take a photo*, and were produced by a participant in chain 4, generation 5, and show use of the same thematic marker, a camera gesture.

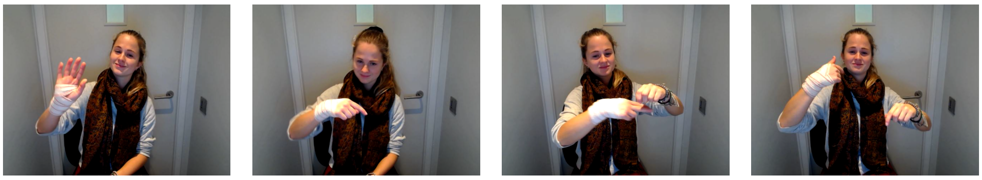
thematic dimension. Figure 2.7b illustrates a gesture for *photographer* in a generation 5 participant, compared with further generation 5 gestures for *camera* and *to take a photo*, where all meanings share the semantic category of photography. The camera shape is used across this thematic category and is the sole signifier of the thematic category, paired in generation 5 with either a point-at-self for *photographer* or a point-at-object for *camera*, which signal the functional category (person or object).

Similar processes of re-analysis and re-combination are widespread by generation 5. The example shown in figure 2.8 illustrates reanalysis from a highly iconic pantomime to a symbolic grammatical marker, demonstrating that the combinatorial systems that emerge by generation 5 are a product of cumulative reanalysis. In generation 1, the gesture for *hairdresser* involves a pantomime in which the hairdresser waves to the customer, motions them to sit down, and mimes cutting hair. The gesturer finally points to herself, an indication of the person category. This hand wave gesture is repeated for the same meaning at generation 2, and by generation 3 has spread to other meanings within the same thematic dimension. By generation 5, the same element has been reanalysed as a functional category marker, re-used throughout the functional category for *action*. Its change in form attests to this; through increasingly restrained movement and a decoupling from directed eye-gaze and facial expression, it is no longer an iconic representation of a greeting, but a systematic, symbolic marker.

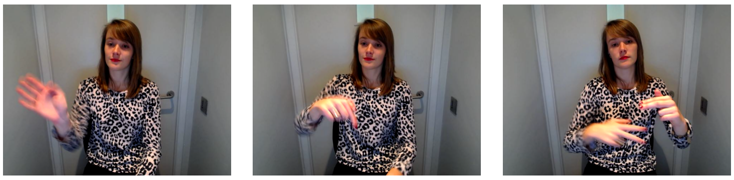
2.3.3 Results: Quantitative results

We investigate the effects of interaction and transmission on the emergence of structure in a manual communication system, and predict that the presence of both mechanisms will lead to gestures that are both efficient and systematic. We measure the efficiency of gestures using a combination of gesture length and the number of repetitions within a gesture. We also measure the entropy of the gestures for each participant. Entropy points to the internal consistency of the sets; do participants use idiosyncratic gestures to describe each meaning, like pantomimes, or do they re-use gestures across meanings, taken from a limited set? Finally, we use the presence of marking on the functional dimension (for the categories *person*, *location*, *object* and *action*) as a proxy for the presence of systematic structure. We expect that gestures produced over five generations will show shorter length and reduced repetitions, demonstrating greater efficiency. We predict that participants will re-use gestures from an increasingly limited pool, and that gestures will be re-used systematically according to the dimensions of the meaning space.

Gesture sequences for each meaning were coded by the first author for shape and

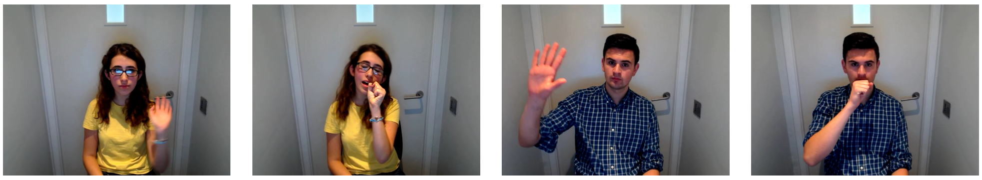


(a) generation 1 gesture for *hairdresser*



(b) generation 2 gesture for *hairdresser*

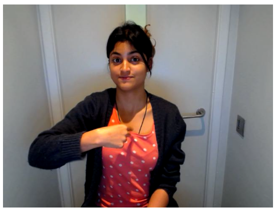
(c) generation 5 gestures for *to give a haircut*



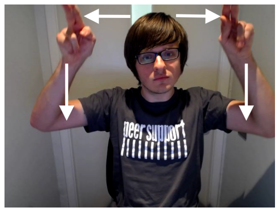
(d) generation 5 gestures for *to sing*

Figure 2.8: Reanalysis of 'wave' pantomime as a functional category marker. The wave gesture starts out as an iconic depiction of a hairdresser's interaction with a customer (a). It is maintained in generation 2(b), and used for thematically related meanings in generations 3 and 4 (not shown). By generation 5 (e,f), the gesture has been reanalysed as a marker for the action category.

handedness. The shape parameter gives a characterising label for the shape of each element that made up a whole gesture, such as *Thumb*, *Book*, or *Box*. Relevant elements such as direction were added to a shape, to distinguish, for example, between a point at an object (point-at-object) and a point at the gesturer's body (point-at-self). The handedness parameter was based on whether the each element of the gesture was one-handed or two-handed. The two parameters for each element were combined and an array of coded elements provided a description of each gesture. For example, [2hBook, 1hPoint-at-object] would describe a gesture in which a participant gestures the shape of an open book with two flat palms and then points at the book object with a one-handed point. Gestures were also coded for the presence of marking along the functional dimension, where a marker is defined as any part of signal meaningful to the entire functional category, and not just to the particular item in that category. The examples shown in figure 2.9 are typical markers for their categories.



(a) marker for person category



(b) marker for location category

(c) marker for object category

(d) marker for action category

Figure 2.9: Examples of category markers used to distinguish items in different categories of the functional dimension of the meaning space. Examples are given from each category, for *person* (a), *location* (b), *object* (c) and *action* (d)

Efficiency

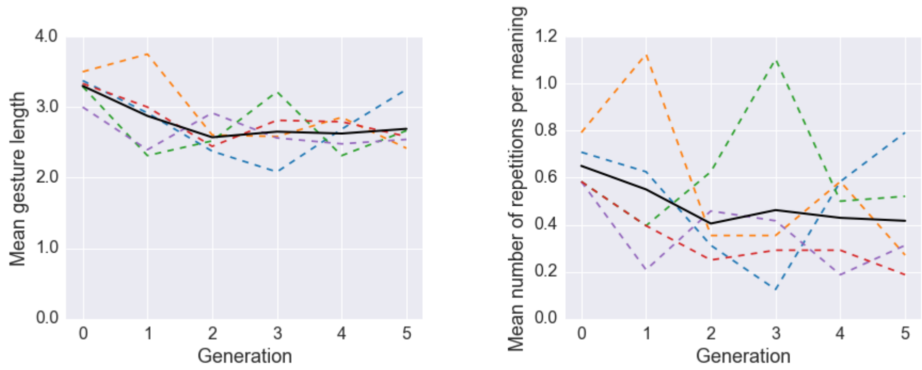
The most straightforward measure of efficiency is perhaps gesture length; all things being equal, shorter gestures encoding a given meaning are more economical. Gesture length, shown in figure 2.10a (left), was calculated as the number of individual elements coded for a particular meaning (e.g. the [2h-book, 1h-point-at-book] gesture described above would have a length of 2). We investigated change in the length of gestures over generations using a linear mixed effects analysis, including generation as a fixed effect. Chain, participant and meaning were included as random effects. We included random intercepts for all random effects, as well as a by-chain random slope of generation. The random effects for participant were nested within chain.[3] The model demonstrated improved fit over the null model ($\chi^2 = 8.68, p = 0.003$) and showed a significant effect of generation ($\beta = -0.16, SE = 0.04, t = -3.54, p < 0.001$). As gestures are transmitted over generations, the length of gestures reduces, though this effect appears to slow down from generation 2 onwards.

Interestingly, visual inspection of gestures revealed that in some cases longer gestures were the result of repetitions. As repetitions are informationally redundant, they indicate a particular source of inefficiency in gestures. We examined the frequency of repetitions within the gestures a participant produced, illustrated in figure 2.10a (right). We count as a repetition as any individual gesture (as defined by the coding scheme) that is repeated within the gesture sequence for a particular meaning. For example, a gesture in which the participant points to their own body, mimes taking a photograph and repeats the pointing gesture would have 1 repetition. However, gestures which involve repeated movement (for example, a camera gesture where the movement of pressing the shutter-release is repeated) are not counted as repetitions, as repeated movement may just be a feature of the gesture. The effect of generation on the proportion of repetitions was analysed using a mixed effects model, including generation as a fixed effect. Analysis of the model revealed an improved fit over the null model ($\chi^2 = 4.56, p = 0.03$) and demonstrated a significant effect of generation ($\beta = -0.16, SE = 0.04, t = -3.54, p < 0.001$). As the participants use and transmit gestures, the proportion of repetitions within gestures decreases, demonstrating decreased redundancy within gesture sets over generations.

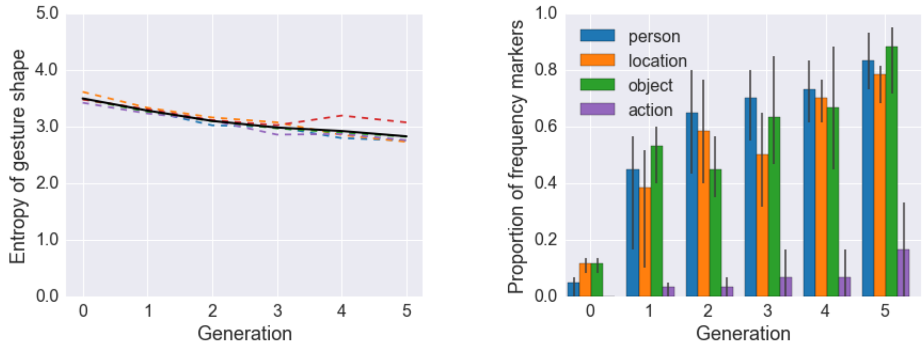
Systematicity

To investigate the consistency of the systems participants produced, we measured entropy of each set of gestures. For example, if a participant uses a pray gesture for *church*, and the same participant re-uses that gesture for *vicar*, *bible* and *to preach*, they show greater consistency than if gestures for those meanings bore no resemblance to one another, and would show low entropy. We calculated the entropy of gesture sets used at each generation, based on the codes for individual gestures described above. Productions from a participant were pooled and entropy (H) was calculated over the

[3] All analyses here and henceforth were implemented using R (R Core Development Team, 2008) and *lme4* (Bates, Maechler, Bolker & Walker, 2015). All models were run using bound optimisation by quadratic approximation (*bobyqa*). Significance values were obtained according to Cunnings (2012). For all analyses in this section, the same random effects structure is used, unless specified.



(a)



(b)

Figure 2.10: (a) Efficiency measures of mean gesture length (left) and mean number of repetitions in a gesture (right), shown for each chain (dashed lines) and averaged across all chains (solid black line). Gestures become more efficient by becoming shorter and including fewer repetitions. (b) Systematicity measures. Mean entropy of gesture shape (left) is shown for each chain (dashed lines) and averaged across all chains (solid black line). Frequency of functional marking (right) shows frequency for each category at each generation, with each coloured bar representing a corresponding functional cateogory. Error bars represent bootstrapped 95% confidence intervals, across 5 transmission chains. Gestures become more systematic and consistent over the set a participant produces. Functional markers accumulate over generations in each category, though gestures for actions generally remain unmarked.

atomic gestures produced (i.e., not gesture sequences), given as

$$H = -\sum p(x) \log_2 p(x)$$

where entropy is summed over unique gestures in a set of gestures. For generation 0, the entropy was calculated over the seed sets used for that chain . We performed a linear mixed effects analysis, analysing the relationship between generation and entropy of the gesture sets. We included generation as a fixed effect, with chain included as a random effect with a random intercept and a random slope of generation. Our model represented an improved fit over the null model ($\chi^2 = 13.02, p < 0.001$) and exhibited a significant effect of generation ($\beta = -0.12, SE = 0.02, t = -7.08, p < 0.001$). Our results indicate a reduction in entropy over the course of the experiment, illustrated in figure 2.10b (left)[4]. As the systems are transmitted, the sets of gestures participants use become more consistent and less variable, with participants in later generations using fewer individual gestures in higher frequencies, rather than a large number of independent gestures.

To investigate the use of systematic markers on the functional category, we counted the frequency of markers used in each generation of each chain, for each category in the functional dimension (*person*, *location*, *object* and *action*; see figure 2.10b (right)). Category types were collapsed into either nouns (*person, location, object*) or verbs (*action*), to examine the emergence of broad functional categories. We performed a logistic mixed effects model to assess the relationship between generation, category type (noun or verb) and the presence of functional marking. Generation, category type and their interaction were included as fixed effects, including random intercepts for chain, participant and target meaning, as well as a by-chain random slope for the effect of generation, and a by-participant random slope for category type. Analysis of the model represented an improved fit over the null model ($\chi^2 = 37.81, p < 0.001$) and revealed a significant effect of generation ($\beta = 0.71, SE = 0.11, z = 6.64, p < 0.001$) and category type ($\beta = -3.5, SE = 0.92, z = -3.84, p < 0.001$). We found no significant effect of the interaction between generation and category type on the presence of functional marking ($\beta = -0.18, SE = 0.18, z = -1.05, p = 0.29$). Participants introduce marking for functional categories, and use of such marking increases as the systems are transmitted. Furthermore, participants create subdivisions within the functional dimension; noun categories carry functional marking, whilst verbs generally remain unmarked (though the presence of verb marking does increase over

[4]Though the trend consistency across chains appears remarkable, further analysis suggests that this consistency is due to common strategies participants use in the task. It is therefore possible for the underlying distributions in each chain to be different from each other and still produce similar entropy values. See Supplementary Materials (section 2.6) for more details.

generations).

2.3.4 Discussion

The results from experiment 1 demonstrate the evolution of systematic, efficient signals in a gestural communication system, emerging from largely unstructured pantomime. The gestures produced by participants show a movement from holistic, memetic structure to segmentable, more language-like structure. Gestures produced by our participants become more efficient; the length of gestures reduces over generations and gestures show fewer repetitions, suggesting that participants produce signals that require less effort and contain distinguishing information necessary for quick and accurate communication. Gestures also become more systematic. Participants increasingly re-use gestures from a more limited set, instead of creating different gestures to describe individual meanings, and use those gestures systematically to distinguish between items in the meaning space. By generation 5, gestures are no longer complex, independent pantomimes, but a system of segmented, interdependent signals. A widespread marking system, which comprehensively covers the meaning space, emerges gradually. Marking along the functional category is introduced early on in participants' gestures, but does not cover the meaning space fully until generation 5. It is not simply the need to communicate these meanings to a partner that lead to the emergence of marking, as the first interacting pairs (or even individuals within a pair) do not mark consistently across categories. Our findings are consistent with previous work investigating the effects of transmission (Kirby et al., 2008; Smith & Wonnacott, 2010) and interaction (Fay & Ellison, 2013; Theisen et al., 2010), as well as both mechanisms in combination (Kirby et al., 2014; Kirby et al., 2015). We suggest that the gestures that emerge in these systems are driven by competing pressures for systems to be both learnable and communicatively efficient, mirroring processes that occur in natural languages (Kemp & Regier, 2012; Regier, Kemp & Kay, 2015). Furthermore, these results are consistent with the emergence of structure in new sign languages. Emerging sign languages arise when a community of deaf individuals come together, and become more systematically structured through repeated use and transmission to new learners (Goldin-Meadow et al., 2014; A. Senghas et al., 2004).

Although these results therefore suggest that both interaction and transmission play a role in shaping the gesture systems produced by our participants, it is unclear exactly to what extent the changes attested to here are the product of either transmission, interaction, or a combination of the two. Experiment 2 attempts to further understand the effects of transmission and interaction by looking at the roles each may play in isolation, with participants taking part in either a transmission-only condition (where individual participants produce gestures without interacting, and those gestures are transmitted to a new learner) or an interaction-only condition (where two participants communicate repeatedly within a pair, without transmission to new learners). Based on previous results (Carr et al. 2016; Kirby et al. 2008; Kirby et al. 2015), we predict that transmission alone will lead to learnable, systematic signals that lack communicative efficiency. The re-use and re-combination of signals provides a compressible set of gestures for a naive participant to learn, but efficiency in production will not necessarily be required. Conversely, interaction alone will lead to idiosyncratic signals suited to quick and effective communication between participants, but which may be unsuitable for learning by naive users, lacking the systematic structure found in systems that have undergone transmission. In either condition we expect that the gestures participants produce will not exhibit the combination of language-like properties found in experiment 1, thereby demonstrating that only the combined pressures of transmission and interaction will lead to linguistic structure; neither pressure alone is sufficient.

2.4 Experiment 2: isolating transmission and interaction

2.4.1 Methods: experiment 2

Participants

35 participants we recruited from the University of Edinburgh careers website to take part in an interaction-only condition, and a transmission-only condition. 10 participants were recruited first for the interaction-only condition (6 female, 4 male, aged 21 to 35, median age 22), followed by 25 participants for the transmission-only condition (17 female, 8 male, aged 19 to 31, median age 23). Random assignment was not used, as the vastly different running times of the two conditions (approximately 90 minutes for the interaction-only condition and 45 minutes for the transmission-only condition) meant that participants were paid different amounts for participation and had to commit to experiments of different lengths. All participants were right-handed native English speakers, with no knowledge of sign language. Participants in the interaction-only condition were compensated £10 for participation, and participants in the transmission-only condition were compensated £5.

Materials

All materials used in the two conditions were the same as used for experiment 1.

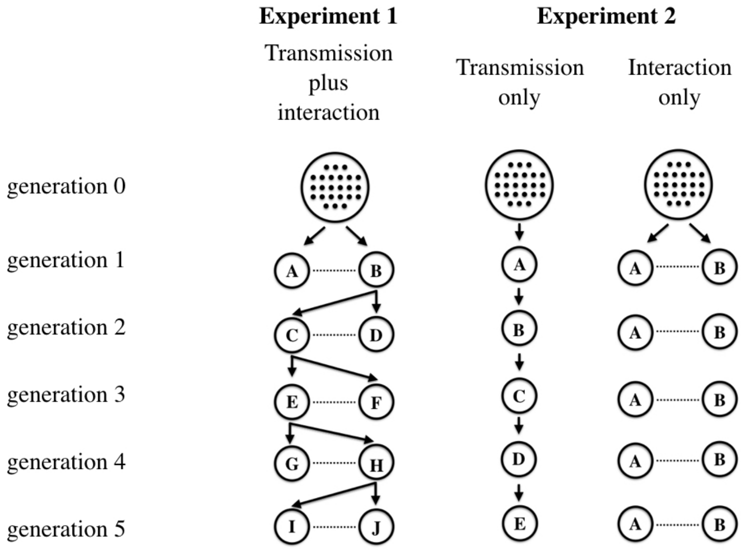


Figure 2.11: Transmission chain structure for all conditions in the two experiments. Solid arrows represent transmission, dashed lines represent interaction. The seed gestures (generation 0) act as the variable starting point for each chain or pair in each condition. Participants in the transmission-only condition only learn from previous participants and pass gestures on to new participants; they do not interact with other participants. Participants in the interaction-only condition only interact and do not pass on their gestures to new participants.

Procedure

The procedure for the two conditions remained largely the same as that described in section 2.3.1 for experiment 1. Participants in both conditions were instructed to use only gesture, not to speak or use manual spelling, and to remain seated throughout the task. While recording, participants were shown themselves on-screen, with the image mirrored, as in experiment 1. Figure 2.11 illustrates the structure of each condition, compared with experiment 1.

Transmission-only condition Participants in the transmission-only condition learnt from previous participants and then had to produce their own gestures for the same meanings. The 25 participants in this condition took part individually, and were not involved in communicative interaction. Participants were organised into 5 transmission chains of 5 generations, with a single participant at each generation (see figure 2.11).

The training stage operated as in experiment 1, where participants were shown videos of another person gesturing, from which they had to guess the meaning the person in the video was trying to communicate. Participants were given feedback, asked to watch the video again in full, before recording their own version of the gesture to camera. Participants in the first generation were exposed to a set of videos made up from the seed gestures, and participants in subsequent generations were shown a subset of videos recorded in testing by the participant from the previous generation. As in experiment 1, participants were shown gestures for 18 out of the 24 items in the meaning space.

In the testing stage, participants were presented at each trial with an item from the meaning space and asked to communicate that meaning to camera using only gesture. A 3 second countdown prepared participants for the start of recording, and once they had finished their gesture, participants could stop the recording by pressing space bar. They completed 24 testing trials, covering all items in the meanings space. Participants in the transmission-only condition were not offered a bonus for fast and accurate responses, unlike in experiment 1 and the interaction-only condition in experiment 2, and were not explicitly timed, in order to remove pressures associated with communication and ensure that gestural output could be attributed only to pressures associated with learning.

Interaction-only condition Participants in the interaction-only condition took part in pairs, repeatedly interacting with each other without transmission to new participants. The 10 participants in the interaction-only condition were organised into 5 pairs. Each pair took part in an initial training round, identical to the training that generation 1 participants undertook in experiment 1. Individually, participants were presented with videos of another person gesturing one of the meanings from the meaning space, and had to identify the correct meaning of the gesture. They were given feedback, asked to watch the video again in full, and then asked to copy the gesture to camera. The training sets were made up of videos from the seed set of gestures, one of the two videos for each meaning selected at random to make a full set of 24 gesture videos. As in experiment 1, participants were shown a subset of gestures for 18 out of the total 24 items in the meaning space, with the order of presentation randomised.

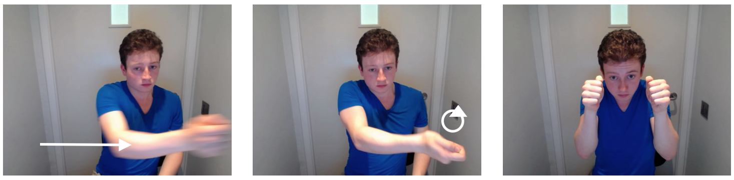
The testing stage in the interaction-only condition was also identical to the testing in experiment 1. Participants had to communicate in pairs, taking it in turns to either communicate (as director) or interpret (as matcher) for all items in the meaning space. However, each pair took part in 5 consecutive testing rounds for the remainder of the experiment, parallel to the 5 generations of experiment 1 (see figure 2.11). As in experiment 1, participants were offered a bonus cash prize for the pair with the highest score, judged as a combination of speed and accurate interpretation of gestures.

2.4.2 Results: qualitative results

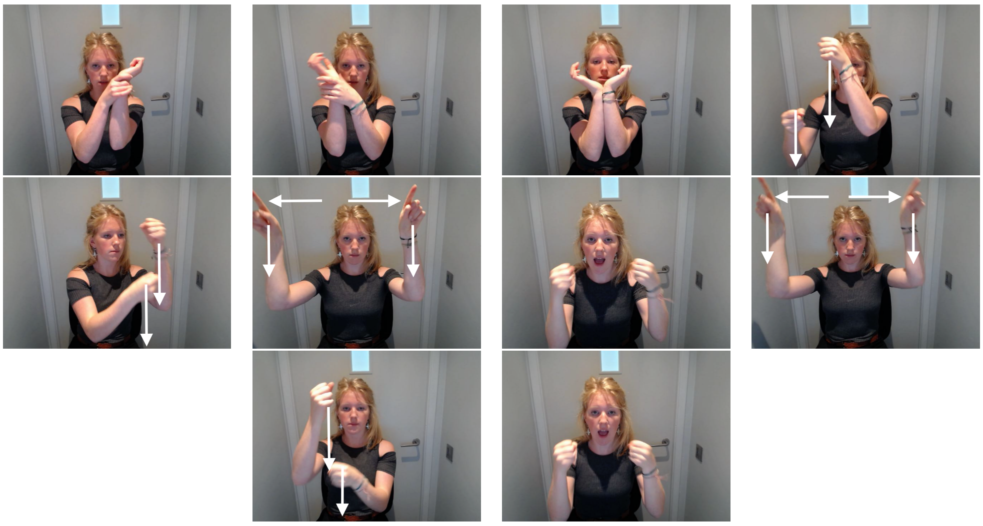
To illustrate the differences between gestures in the transmission-only and interaction-only conditions, and to contrast them to the gestures elicited in experiment 1, we present typical examples from each condition, showing the evolution of these gestures as they are used repeatedly or transmitted to new learners.

Gestures in the transmission-only condition Similar to gestures in experiment 1, gestures in the transmission-only condition show some use of marking, which emerges over generations. Figure 2.12 shows gestures for *prison* for a generation 1 participant and a generation 5 participant of the same chain. In generation 1, the gesture is holistic, lacking any structure that is systematically re-used across the set of gestures; by generation 5, some systematicity emerges in the form of a location marker (the 'box' gesture), contrasted with the rest of the gesture sequence signalling the thematic category. However, the generation 5 gestures lack the clear two-part structure found in experiment 1. Repetition and redundancy are common and there is an asymmetry between the number of components used to signal the functional dimension and those used to signal the thematic dimension. Similar properties are present in the same participant's gesture for *church*, shown in figure 2.12b. She consistently uses a location marker (though the roof-like properties are more noticeable here) combined with multiple signals for the thematic dimension of religion (using and repeating both the cross and the bible gesture). We also see the re-use of particular signals to convey the thematic dimension, such as that shown in figure 2.13 for *photographer*. The participant here also uses a number of other redundant signals that are not systematically used across either thematic or functional categories. For example, in figure 2.13, the participant produces a sign in the middle of the gesture where she looks at the photographs she has taken (middle row, panels 2 and 3), presumably on a digital camera. This gesture is not re-used in her gesture for other gestures in that thematic category and seems largely redundant for distinguishing the meaning from others, given the point-at-self that begins and ends the gesture, differentiating it from *camera* and other meanings.

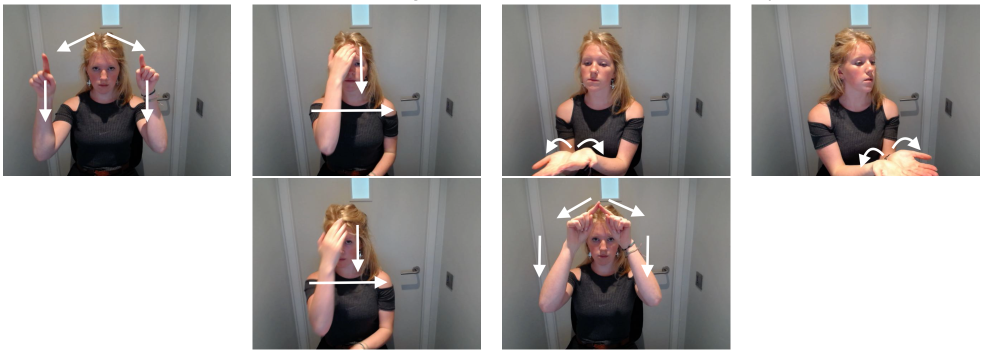
Gestures in the interaction-only condition In the interaction-only condition, widespread systematic structure is not present in the gesture sets participants produce. Figure 2.14 illustrates the development of gestures for *prison* from round 1 to round 5 for a single participant, in comparison with his gesture for *church* at round 5. The



(a) Gesture for *prison* from generation 1, transmission-only condition.



(b) Gesture for *prison* from generation 5, transmission-only condition.



(c) Gesture for *church* from a generation 5, transmission-only condition.

Figure 2.12: Emergence of functional marking over generations in the transmission-only condition. Gestures are for the meaning *prison*, taken from generation 1 (a) and generation 5 (b) of the same chain, compared with the gesture for *church*, produced by the same generation 5 participant (c).

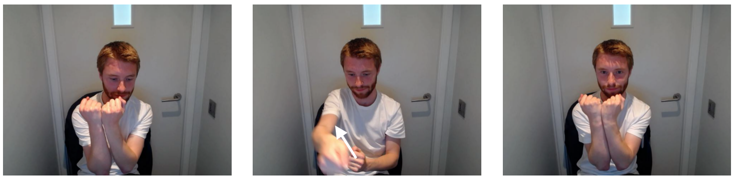
Figure 2.13: Redundancy in generation 5 of the transmission-only condition. The entire gesture sequence shown here communicates *photographer*.

participant uses no segmentation or marking to distinguish the dimension of the meaning space, and, by round 5, the gesture for *prison* has changed to a short gesture, idiosyncratic for that particular meaning, demonstrating efficiency but a lack of structure. When viewed alongside his gesture for *church*, it is evident that the two gestures share little that would signal the shared category of location. The two gestures, despite sharing the functional meaning of location, are independent of each other.

Gestures produced by participants in the interaction-only condition also do not systematically re-use the same thematic signals across thematic categories, but instead use gestures that demonstrate different salient properties of the thematic category. For example, figure 2.15 shows gestures from a participant in the interaction-only condition, in round 1 and round 5. In a round 1 gesture for *photographer*, the participant performs a point-at-self followed by a camera outline. This is recognisable in the same gesture in round 5, though she signals both components at the same time, reducing the camera outline to a one-handed half-outline. In comparison, her round 5 gesture for *camera* does not use the camera outline at all, instead detailing the camera lens, rather than the main body of the camera. Gestures in this condition show notable reduction in form, but again lack the systematic structure found in gestures produced in experiment 1.

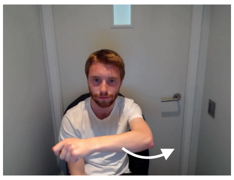
Quantitative analysis

Analyses presented below provide results examining both the transmission-only and interaction-only conditions, in comparison with the results from experiment 1. We



(a) Gesture for *prison* from a participant at round 1.

(b) Gesture for *prison* from a participant at round 5.



(c) Gesture for *church* from a participant at round 5.

Figure 2.14: Lack of segmentation and marking of the meaning space dimensions across rounds in the interaction-only condition. Gesture shown are for *prison*, from round 1 (a) and 5 (b) for the same participant in the interaction-only condition, compared with a gesture from the same participant at round 5, for the meaning *church*.

refer to experiment 1 here as transmission-plus-interaction for consistency and clarity.[5]

Efficiency

To investigate efficiency across conditions, we noted gesture length and frequency of repetitions, as in experiment 1 (see section 2.3.3). A linear mixed effects model analysed the effect of generation on the length of gestures in each condition, in comparison with gesture lengths in experiment 1. Generation (or round), condition and their interaction were included as fixed effects, with transmission-plus-interaction (experiment 1) used as the baseline condition. Chain (or pair), participant and target meaning were included in the random effects structure, all with random intercepts. A by-chain random slope of generation was included, and the effects structure for participant was nested within chain, or pair [6]. Analysis of the model showed a significant improvement over the null model ($\chi^2 = 198.67, p < 0.001$) and revealed a signific-

[5]See Supplementary Materials for additional results pertaining to matching accuracy, alignment between participants, learnability (measured as transmission success), and gesture order.

[6]All models use transmission-plus-interaction (experiment 1) as the baseline condition. In the models, chain is analogous to each pair in the interaction-only condition and generation is analogous to each round that a pair takes part in

(a) Gesture for *photographer* from a participant at generation 1.

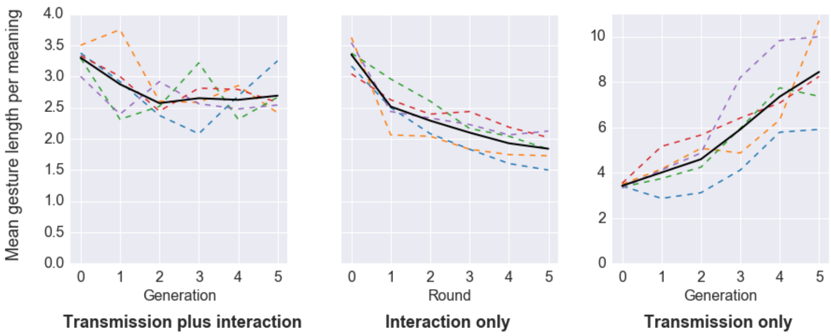
(b) Gesture for *photographer* from a participant at generation 5.

(c) Gesture for *camera* from a participant at generation 5.

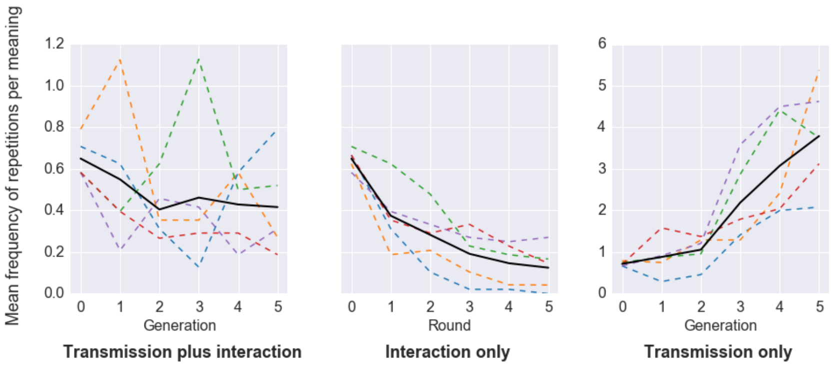
Figure 2.15: Gestures in the interaction-only condition highlight salient category features rather than systematic marking across categories. Gestures show *photographer* at round 1 (a) and 5 (b) for the same participant in the interaction-only condition, compared with a gesture from the same participant at round 5, for the meaning *camera*.

ant effect of generation ($\beta = -0.14, SE = 0.06, t = -2.33, p = 0.02$). There was no significant interaction found between generation and the interaction-only condition ($\beta = -0.05, SE = 0.06, t = 0.86, p = 0.39$), but the model revealed a significant interaction of generation and the transmission-only condition ($\beta = 1.11, SE = 0.08, t = 13.50, p < 0.001$). Gestures in the transmission-plus-interaction condition and the interaction-only condition show a reduction in length over the course of 5 generations/rounds, whereas gestures in the transmission-only condition show an increase in gesture length over generations (see figure 2.16a). This may, in part, be explained by the absence of a time pressure in the transmission only condition, which was present in the other two conditions.

To assess redundancy in the gestures participants produced, we noted the number of repetitions within a single gesture sequence, in each condition (as described in section 2.3.3). Figure 2.16b illustrates this measure, showing the mean number of repetitions within a single gesture, for each condition. A linear mixed effects analysis assessed the fixed effects of generation or round, condition and their interaction. The model represented an improved fit over the null model ($\chi^2 = 191.37, p < 0.001$). The model did not reveal a significant effect of generation ($\beta = -0.04, SE = 0.03, t = -1.312, p = 0.09$), and revealed no significant interaction between generation and the interaction-only condition ($\beta = -0.02, SE = 0.04, t = -0.75, p = 0.24$). The model



(a)



(b)

Figure 2.16: Gesture length (a) and frequency of repetitions (b) for all conditions. Over generations, gestures become shorter, with fewer repetitions only in the conditions with interaction. When only transmission is involved, the reverse trend is observed. Dashed lines represent data in each chain/pair, solid black lines represent the average across all chains/pairs. Note the different y-axis scales for the transmission-only condition.

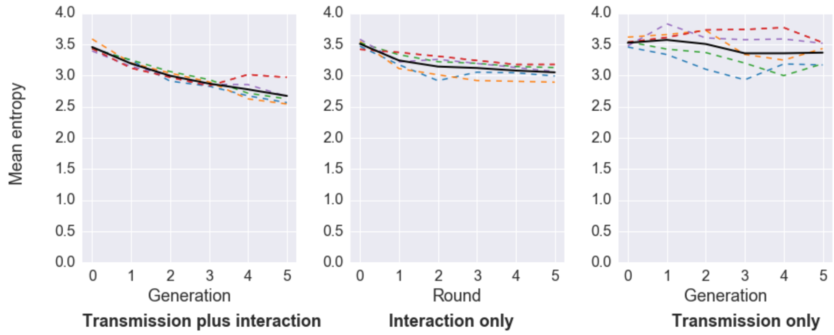
did demonstrate a significant interaction between generation and condition for the transmission only condition ($\beta = 0.63, SE = 0.05, t = 12.82, p < 0.001$). Gestures in the transmission-plus-interaction condition and the interaction-only condition do not demonstrate an increase in redundancy over generations, and suggest a decrease in redundancy. Gestures in the transmission-only condition show a cumulative increase in the number of repetitions used over generations.

2.4.3 Systematicity

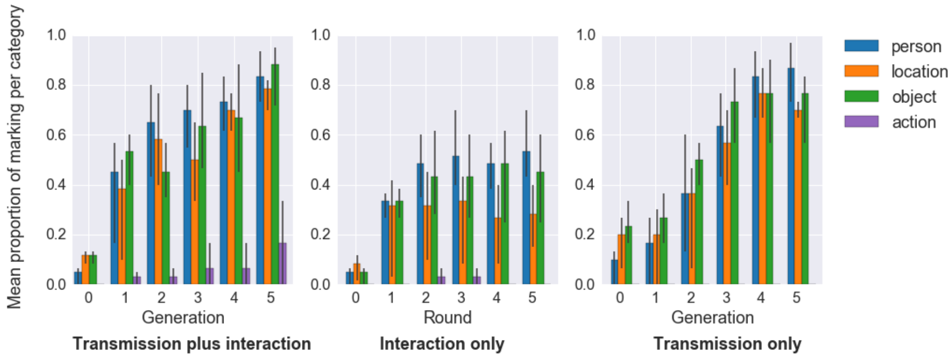
Entropy of the gestures produced for each meaning was calculated for both the transmission-only and interaction-only conditions, using the same procedure described in 2.3.3. Figure 2.17a shows the average entropy for all chains and pairs in each condition.

A linear mixed effects analysis analysed the effect of generation on gesture shape entropy in each condition, using the transmission-plus-interaction condition (experiment 1) as the baseline condition. Chain (or pair) and participant were included as random effects, along with a by-chain random slope of generation. Random effects for participant were nested within chain. Our model represented an improved fit over the null model ($\chi^2 = 118.82, p < 0.001$) and showed a significant effect of generation for the baseline condition ($\beta = -0.12, SE = 0.01, t = -10.27, p < 0.001$). The model showed a significant effect of generation in the transmission-only condition ($\beta = 0.16, SE = 0.06, t = 2.69, p < 0.009$), as well as a significant interaction between generation and condition ($\beta = 0.08, SE = 0.02, t = 4.15, p < 0.001$). The model revealed a significant interaction between generation and the interaction-only condition ($\beta = 0.08, SE = 0.01, t = 5.57, p < 0.001$), though no effect of generation was found for the interaction-only condition ($\beta = 0.007, SE = 0.05, t = 0.13, p = 0.89$). Gesture shape entropy decreases over generation in the transmission plus interaction condition. Entropy in the interaction-only and transmission-only conditions does not reduce to the same extent (though there was still a significant decrease in entropy in these conditions, as indicated by significant effects of generation in regression analyses using the interaction-only and transmission-only conditions as baseline).

As in experiment 1, we examined the frequency of markers for categories in the functional dimension for the categories *person*, *location*, *object* and *action*. Frequencies were noted across both the transmission-only and interaction-only conditions, shown in comparison with experiment 1 in figure 2.17b. A logistic mixed effects analysis examined the effect of generation and condition on the presence of categorical markers in the gestures participants produced. Our model incorporated generation and condition as fixed effects, as well as their interaction. Analysis of the model revealed an im-



(a)



(b)

Figure 2.17: (a) Mean entropy of gesture sets in each condition. Solid black lines represent the mean across all chains/pairs, coloured dashed lines represent mean values for each chain/pair. Entropy reduces across all conditions, but reduces to a greater extent in the condition in which both interaction and transmission are at work. (b) Frequency of functional marking at each condition, shown at each generation for the categories of *person*, *location*, *object*, *action*. Error bars represent 95% bootstrapped confidence intervals. Strikingly, the frequency of functional markers only increases in conditions with transmission to new learners. Gestures in the interaction-only condition do not show an increase in functional marker frequency after round 1.

proved fit over the null model ($\chi^2 = 50.89, p < 0.001$). The model showed a significant effect of generation ($\beta = 0.75, SE = 0.09, z = 8.29, p < 0.001$), and a significant interaction between the generation and the interaction-only condition ($\beta = -0.54, SE = 0.09, z = -5.82, p < 0.001$), though no significant interaction was found between generation and the transmission-only condition ($\beta = 0.14, SE = 0.11, z = 1.25, p = 0.21$). Frequency of category marking in the transmission-plus-interaction condition and transmission-only conditions increases cumulatively over the generations of transmission. In the interaction-only condition, pairs introduce some category markers in generation 1, but marker frequency remains lower than in both the transmission-only and transmission-plus-interaction conditions. Indeed, a model fitted to data from generation 1 to 5 (i.e. without seed data), does not show a significant effect of generation for the interaction-only condition ($\beta = 0.12, S = 0.11, z = 1.13, p = 0.26$), confirming that gestures in the interaction-only condition do not show an increase in the frequency of functional markers after generation 1.

2.4.4 General discussion

We investigated the roles that interaction (using a system to communicate) and transmission (to new learners) play in shaping a language. We set out to understand how these processes facilitate the continuous evolution of pantomimes into language-like gestures, involving a shift from holistically structure, inefficient gestures to ones exhibiting sysytematicity and efficiency.

Our results provide a clear picture of how two key mechanisms– transmission and interaction– interact in the emergence of a communication system. Experiment 1 showed a combination of interaction and transmission leading to the emergence of language-like systems from pantomime. In experiment 1, participants showed increased efficiency through the use of shorter gestures and a drop in redundancy. Gestures in experiment 1 gradually became more systematic in structure, showing the creation of a system of interdependent parts that were re-used across meanings.

Experiment 2 looked at each mechanism in isolation, and found that, although gestures in each condition demonstrated an increase in some of the properties we associate with linguistic systems, they did not show the emergence of language-like properties to the same extent as the gestures in experiment 1. Gestures in the interaction-only condition showed increased efficiency, and consistency. However, the signals used to mark categories did not spread throughout the systems; gestures did not become fully segmented and instead became increasingly idiosyncratic. The sets of gestures produced here are consistent with previous studies, where interaction between communicators led to a cumulative increase in communicative efficiency

(Fay & Ellison, 2013; Fay et al., 2010; Garrod et al., 2007; Theisen et al., 2010). Signals in these cases lack systematic structure, and individual signals remain relatively independent of each other. Without new learners being introduced, there is little pressure for learnability; gestures used within a pair do not need to be predictable across the meaning space for communicative success.

The transmission-only systems, on the other hand, showed widespread use of marking to signal the functional dimension, comparable with experiment 1 and demonstrated segmentation and increasingly systematic structure. The present findings are consistent with iterated learning studies, demonstrating cumulative increase in structure as signals are transmitted to new learners (Kirby et al., 2008; Verhoef et al., 2014). Naive participants were able to learn and reproduce the gestures they saw in training, and re-used and recombined signals into a system of interdependent gestures. However, without the pressure for efficient communication, gestures produced by individuals in the transmission-only condition were much longer and exhibited large-scale redundancies through repetition.

The set of experiments described here demonstrate the effects of transmission and interaction on a linguistic system. Importantly, neither transmission alone nor interaction alone led to language-like structures. Only when both mechanisms worked together did we see the emergence of the segmentation and conventionalisation associated with linguistic structure. The miniature artificial sign languages that evolve in these experiments show evidence of adapting to the specific pressures at play in each condition; systems in the interaction-only condition become suited to efficient communication within a pair, whilst systems in the transmission-only condition lack communicative efficiency, but demonstrate systematic structure which signals the dimensions of the meaning space. These results are corroborated by natural data from emerging sign systems. Homesigns, used by individuals, lack regularity and exhibit low rates of conventionalisation (Goldin-Meadow et al., 2014; Richie et al., 2014). Emerging sign languages, in their earliest stages, begin to show stabilisation and conventionalisation but lack the consistency of older sign languages; as the systems are used in interaction and transmitted to new learners, the languages further stabilise and begin to develop consistent and regular structures across signers (Goldin-Meadow et al., 2014; Padden et al., 2010; Sandler et al., 2005). Our results demonstrate similar patterns; gestures in early generations (generations 0 and 1) show lack of regularity and conventionalisation. Gestures that are used between pairs of participants become more conventionalised and efficient, but lack systematic structure. It is only through the repeated use and transmission of the systems that gestures become systematic and regular within a chain. In particular, our results show the development of categorical markers that distinguish between nouns and verbs in the meaning space, consistent with early development of such categories in emerging sign languages (Goldin-Meadow et al., 2014; Padden et al., 2013; Tkachman & Sandler, 2013). Furthermore, our results exhibit a pattern consistent with the results of Goldin-Meadow et al. (2014), where stability of categorical distinctions increased over cohorts of NSL from Nicaraguan homesigners (who showed little stability) to second-cohort signers. By using the manual modality, our experiments minimise the interference that prior linguistic knowledge may have in artificial language learning experiments. The present framework allows the investigation of modality-independent mechanisms that affect language, but may also offer a platform for investigating modality-specific phenomena involved in the emergence of linguistic structure, such as the presence of iconicity in a system as it develops structure. Our findings support the hypothesis that pressures for learnable and communicatively effective systems drive the emergence of language-like structure, and that both pressures must be present for the emergence of signals that are both systematic and efficient (Kirby et al., 2015; Regier et al., 2015). The results from experiment 2 support our findings from experiment 1 by demonstrating that neither transmission nor interaction, when isolated, lead to the emergence of language-like systems.

2.5 Conclusion

We set out to investigate how linguistic structure emerges in manual sign systems. By incorporating both interaction within generations and transmission of the system to naive learners in our experiments, we have shown how the interaction of these mechanisms drives the gradual emergence of systematic and efficient gestures. The studies presented here offer a parallel to the observation of structural development in natural sign languages and experimental studies in the spoken and written modalities. Our findings demonstrate the emergence of language-like signs from non-linguistic signals and point to the importance of both transmission and interaction for the emergence of linguistic structure. The experimental framework presented here combines silent gesture and iterated learning, offering a novel combination of these two paradigms; by using the manual modality, our experiments minimise the interference that prior linguistic knowledge may have in artificial language learning experiments. Our findings provide an experimental complement to data from natural languages, that allows us to bridge the gap between longitudinal, naturally occurring data, and constrained, manipulable experimental work.

2.6 Supplementary Analyses

Measures analysing video length in each trial and matching accuracy offer further insight into how the gestures systems change over generations. Similarly, a further investigation of the entropy measure used in experiment 1 (section 2.3.3) is used to understand how gestures change as system entropy reduces. I measured alignment between communicating partners and learnability across generations, further corroborating our main claim that transmission facilitates learnable systems, whilst interaction facilitates systems suited for communication. Finally, I briefly discuss ordering in gesture sequences.

2.6.1 Results

Video length

Recording length for videos in each trial was noted at each generation. Figure 2.18 shows mean video length over generations in each condition.

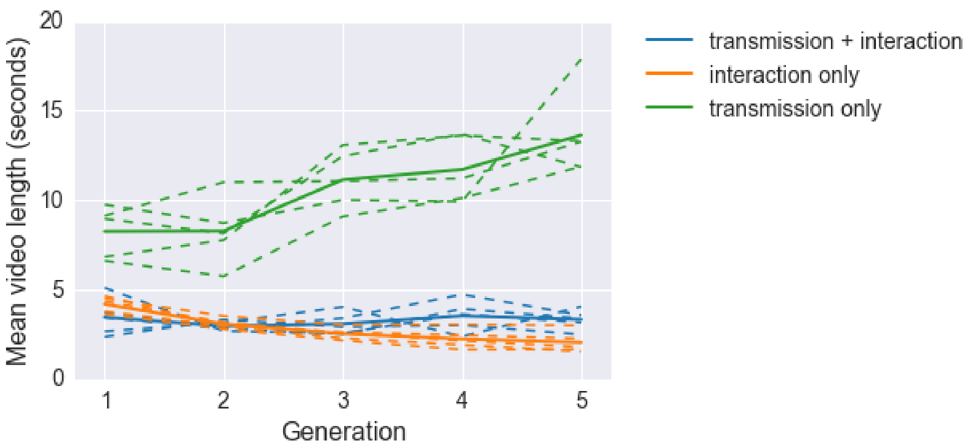


Figure 2.18: Mean video length over generations in each condition. Solid coloured lines represent overall means in each condition, dashed lines represent means in each chain or pair. Video length remains stable in experiment 1 (transmission and interaction), reduces over generation in the interaction only-condition, and increases over generations in the transmission-only condition.

A linear mixed effects model analysed the effect of generation and condition on video length, as well as their interaction. Chain, participant and target item were included as random effects, with a by-chain random slope of generation and the random effects for participant nested within chains. Experiment 1 (transmission and interaction) was used as the baseline condition. The model demonstrated significantly better fit over the null model ($\chi^2 = 221.34, p < 0.001$). Analysis of the model did not reveal a significant effect of generation for the baseline condition ($\beta = 0.03, SE = 0.13, t = 0.26, p = 0.80$). However, the model showed significant interactions between generation and condition both for the interaction-only condition ($\beta = -0.54, SE = 0.12, t = -4.44, p < 0.001$) and the transmission-only condition ($\beta = 1.39, SE = 0.20, t = 6.86, p < 0.001$). Gestures in experiment 1 (transmission plus interaction) show no significant changes in video length over generations. However, gestures in the interaction-only condition reduce in length over generations, whilst gestures in the transmission-only condition increase in length.

Matching accuracy

Matching accuracy was noted in each trial (with a matcher's guess coded either correct or incorrect). For each pair, we calculated the percentage of trials in which the target meaning was correctly identified from the gesture. Table 2.1 gives the percentage of correct trials at each stage (training or testing) in each condition. Across all conditions, participants showed high matching accuracy, with no notable change over generations.

	Transmission plus interaction		**Interaction-only**		**Transmission-only**	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
Training	78.0%	75.3%-80.2%	76.4%	71.4%-80.3%	84.1%	81.4%-86.3%
Testing	87.3%	85.1%-89.0%	94.9%	93.4%-96.0%	-	-

Table 2.1: Mean percentage of correct interpretations made during training and testing in each condition, with 95% confidence intervals. No testing scores are given for the transmission-only condition because no matching trials took place.

Alignment and learnability

We measured the similarity of a participant's gestures to (a) the gestures of their partner in communication, and, (b) the the gestures of their training model. We will call the former within-generation similarity, and the latter between-generation similarity. Gesture similarity is measured using the Jaccard index, given as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{2.1}$$

where the Jaccard index of two sets, A and B is the intersection of the two sets (the elements that the sets share) divided by the union of the two sets (the total of unique

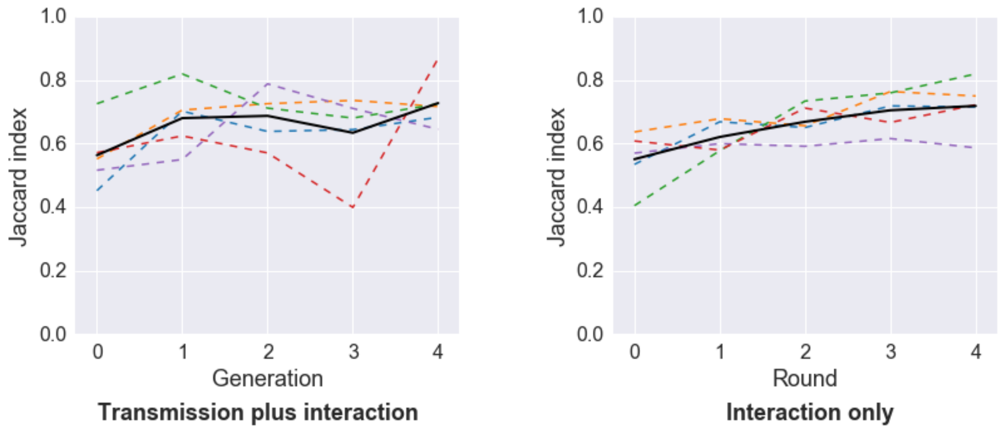
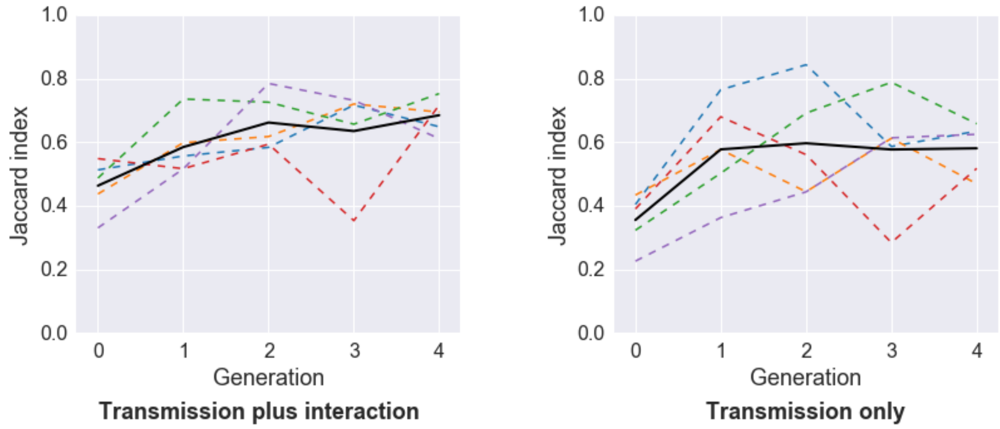


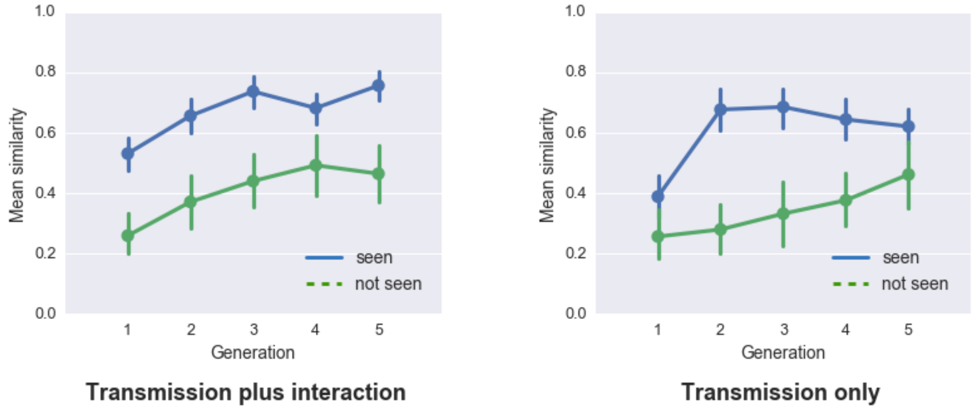
Figure 2.19: Within-generation similarity for experiment 1 (transmission plus interaction) and the interaction-only condition of experiment 2, indicating alignment of communicating pairs. Alignment increases over generation in both conditions, with no significant difference between conditions. Solid black lines represent the mean over all chains/pairs, and dashed coloured lines represent the means for each chain/pair.

elements in both sets). This measure accounts for the elements that are shared in a set, regardless of how they are used, and therefore does not account for the frequency which particular elements occur in each set. The Jaccard index was deemed appropriate for measures of similarity in this case, as this measure is precisely concerned with the gesture elements that appear in each set, rather than their frequencies. Furthermore, it is possible that differences in repetitions of shared elements between participants would obscure the fact that the gestures they use for particular meanings are the same, and the repetition measure described in section 2.3.3.Within-generation similarity was calculated for experiment 1 (transmission plus interaction) and the interaction-only condition in experiment 2, but not for the transmission-only condition (where there is only one participant per generation). Similarly, between-generation similarity was calculated for experiment 1 (transmission plus interaction) and the transmission-only condition in experiment 2, but not the interaction-only condition, as no new learners are introduced, and no further training takes place.

Figure 2.19 illustrates within-generation similarity for the transmission plus interaction and interaction-only results. A linear mixed effects model analysed the effect of generation and condition on alignment. Chain (or pair) and target meaning were included as random effects with random intercepts, and a random slope of generation (or round) was included for chain (or pair). The model showed a significant improvement over the null model ($\chi^2 = 12.16, p < 0.01$). The model revealed a significant effect of generation (round) ($\beta = 0.03, SE = 0.01, t = 2.80, p < 0.01$), but no interaction between generation (round) and condition ($\beta = -0.01, SE = 0.01, t = 0.98, p = 0.33$).



(a)



(b)

Figure 2.20: (a) Between-generation similarity of gestures for experiment 1 (transmission plus interaction) and the transmission-only condition of experiment 2, indicating learnability of gestures over generations. In both conditions, gestures show an increase in learnability; participants in later generations are better at reproducing the gestures they were trained on than participants in earlier generations. Solid black lines represent the mean over all chains, and dashed coloured lines represent means in each chain. (b) Between-generation similarity of gestures based on whether the gestures was seen in training (solid blue line) or not seen in training (dashed green line), for experiment 1 (transmission plus interaction) and the transmission-only condition of experiment 2. Gestures in both conditions show an increase in learnability, and both seen and not-seen gestures demonstrate this increase. An increase in learnability in gestures not seen in training suggests that participants are generalising parts of gestures across meanings, so that new items are identifiable to learners. Error bars represent bootstrapped 95% confidence intervals.

Alignment increases across generations in both the transmission-plus-interaction condition and interaction-only condition, with no significant difference between the two.

We also ran a linear mixed effects analysis for learnability in the transmission-plus-interaction and transmission-only conditions, including generation, condition and target type (i.e. whether or not the target item was seen in training) as fixed effects, as well as their interactions. Chain, participant number and target meaning were included as random effects with random intercepts, and a random slope of generation was included for chain. The model showed an improvement over the null model ($\chi^2 42.84, p < 0.001$). We found a significant main effect of generation ($\beta = 0.05, SE = 0.02, t = 3.00, p = 0.02$), and a significant effect of target type ($\beta = -0.30, SE = 0.05, t = -6.23, p < 0.001$). The model did not reveal a significant difference between conditions ($\beta = -0.07, SE = 0.07, t = -0.95, p = 0.34$). None of the two-way interactions between fixed effects were significant, and no significant three-way interaction between fixed effects was found ($\beta < 0.04$, $p > 0.44$). Participants in both conditions are able to faithfully reproduce the gestures seen in training, and show an increase in faithful reproductions over five generations. Furthermore, although gestures for target meanings not seen in training a reproduced less faithfully, participants are increasingly able to approximate these gestures over generations, suggesting that participants are generalising patterns in gestures they were trained on to gestures for novel meanings.

Entropy

Figure 2.10b (left) illustrates entropy across generations for experiment 1 and suggests a remarkably similar trend in across chains in the experiment. To understand this trend further, figure 2.21 represents the number of gestures against their frequencies for each generation of each chain in experiment one. Unique gesture shapes described by the coding scheme are given on the x-axis, at each generation, and their frequency is shown on the y-axis. These charts demonstrate the similar trajectories that each chain follows in regularising the gesture shapes they use, which is measured in the main text using entropy. We suggest that the ways in which participants create regular gesture sets is limited; they begin with a larger number of different gestures, and through communication and transmission, settle on particular shapes, eliminating competing gestures. This leads to a smaller pool of gestures being reused more frequently, and can occur regardless of the particular gestures used.

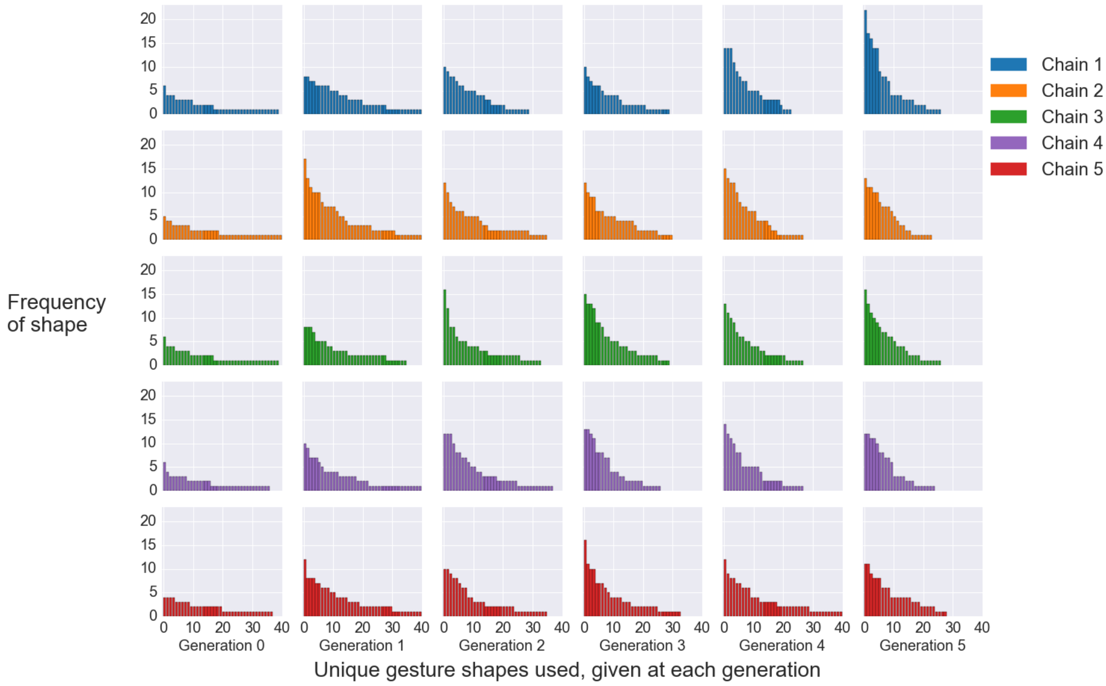


Figure 2.21: Frequency of unique gesture shapes at each generation, for each chain in experiment 1. Over generations in each chain, participants increasingly re-use the same atomic gestures, from a smaller pool of gestures. For example, at generation 0 seed participants use a wide range of different gestures in low frequencies. By generation 5, participants re-use the same gestures in higher frequencies, and use fewer unique gestures.

Gesture order

A frequent application of silent gesture studies is in the study of word order. Silent gesture studies have shown that hearing participants in silent gesture studies forgo the word order of their native language in favour of other orders that are consistent across participants engaged in the task, depending on verbal semantics (Schouwstra & de Swart, 2014), reversibility of the participants in the scene (Hall et al., 2013) and access to a lexicon (Marno et al., 2015), to name a few examples. We analysed the gesture order for gestures where a functional marker was coded as present. Gesture order was simply coded as function-first for gestures where the functional marker was the first element in the sequence, or theme-first, where the thematic signal was the first element in the sequence. Analysis across conditions and generations showed no notable effects; when participants used markers, the order in which they arrange functional and thematic signals was consistent. Across all conditions, participants produce gestures where the functional marker begins the gesture sequence, for the categories of *person*, *location* and *action*. This pattern is reversed, however, for the *object* category, where participants produce theme-first gestures, providing information about the thematic category before the functional category. This reversal is understandable; the most common object marker is the point-at-object gesture, and it therefore seems more sensible for participants to gesture the actual object (the thematic signal) before they point at the object.

	Transmission plus interaction		**Interaction-only**		**Transmission-only**	
	Function first	Theme first	Function first	Theme first	Function first	Theme first
Person	77.7%	22.3%	97.2%	2.8%	51.7%	48.3%
Location	93.2%	6.8%	96.0%	4.0%	89.2%	10.8%
Action	86.4%	13.6%	100%	0.0%	-	-
Object	12.8%	87.2%	11.2%	88.8%	0.0%	100%

Table 2.2: Percentage of gestures in each category produced with the functional marker first (function first) or thematic signal first (theme first). Percentages are given for each condition. There are no ordering data for the *action* gestures in the transmission-only condition, as they did not mark any meanings in this category.

2.7 Discussion

The measures presented here further corroborate our findings throughout this chapter, demonstrating how the presence of interaction and transmission in the construction of a communication system affects the forms present in that system. Gestures in the interaction-only condition reduce in the time they take to produce, suited for rapid communication, whilst gestures in the transmission-only condition become longer over generations. Similar patterns are borne out in the measures shown in figure 2.16, where gesture length measured as gesture units and the number of gesture repetitions decreases in the interaction-only condition, but increases in the transmission-only condition.

Our findings also suggest that the gestural systems participants produce respond to the pressures of transmission and interaction. When communication is present, gestures produced by communicating pairs become more similar; that is, the communicating pair becomes more aligned with one another, consistent with previous work on communicative alignment (Fay et al., 2010; Garrod et al., 2007; Healey et al., 2007). Additionally, when participants learn from a model, that learning process affects the gestures participants produce, leading to gestures that new learners are increasingly able to successfully reproduce, even when they have not been given a model for a target meaning. These results replicate findings across a number of iterated learning experiments, that suggest systems that have undergone transmission become more learnable (Carr et al., 2016; Kirby et al., 2008, 2015).

Close analysis of the frequency of gestures shapes over generations in experiment 1 suggests that systems with differing formal structures show remarkably similar responses to pressures from interaction and transmission, leading in this case to similarly low system entropy, despite the different gestures that are used in each chain.

Finally, analysis of gesture order found no evidence of change over generations, but consistency in gesture order indicates natural ordering tendencies, with participants overwhelmingly producing functional elements before thematic elements to describe people, locations and actions, and producing thematic elements first for gestures in the object category. Although it seems reasonable that object gestures would necessarily put theme first – one must first depict an object to meaningfully point at it –the reversal does not indicate why function-first is preferred overwhelmingly for the other categories. One explanation points to the predictability of elements in a string, where less predictable elements occur first. However, predictability may have multiple interpretations. If predictability is considered in terms of information density (Maurits, Perfors & Navarro, 2010), we would expect to see a preference for theme-first gestures, given the input; there are 6 thematic categories to distinguish between, but only 4 functional categories. An alternative explanation considers predictability in terms of the iconicity of the referent (Givon, 1985), where signals that require more effort to interpret are placed earlier in a sequence. If thematic elements carry more semantic content (i.e. they are more iconic, more specific than functional elements) then they could be considered more accessible to participants and would therefore be placed second to less iconic, more general and less accessible functional elements. A pointing gesture, for example, may be considered more general and less iconic, than a book gesture, a camera gesture or a praying gesture, as a point is meaningful relative to what it is pointing to. However, such interactions between reference and ordering are not trivial, and semantic predictability is just one explanation for producing function-first gestures. Schouwstra (2016) offers an alternative, based on how temporal adverbs are positioned in existing linguistic systems and silent gestures. Hearing participants in a silent gesture task were asked to gesture target items with a temporal element (e.g. ”The dwarf eats the pizza at three o'clock). Participants placed the temporal element first in the sequences of gestures they produced, using the temporal sequence to frame the rest of the proposition. Indeed, in the data collected here, there is often an asymmetry between functional elements and thematic elements, with participants often producing a higher ratio of thematic elements to functional elements; it is possible gesturers use the functional elements as framing for a set of thematic elements. Understanding this phenomenon requires further study of signal ordering in this data set, but the present investigation highlights the utility of silent gesture studies for understanding linguistic phenomena such as element ordering.

2.8 Summary

The work discussed in this chapter presents a novel method to understand the effects of transmission and interaction in manual communication. Our results demonstrate the importance of the combination of the two process in the evolution of language-like structure. These results corroborate findings from previous work on interaction and transmission of linguistic systems. We also put forward a novel method that allows the examination of general linguistic processes, but also modality-specific aspects of gestural communication and silent gesture as an experimental medium. Our experiments minimise the interference of participants' native language, and additionally provide an experimental comparison with emerging sign language that provide crucial insight into early evolving languages. In the following chapters, I will discuss how this method can be extended to bridge the gap between experimental work and field work on such sign languages, as well as how this work can be extended to more complex linguistic phenomena.

Chapter 3

The emergence of linguistic categories: investigating the seeds of the noun/verb distinction

3.1 Introduction

The majority of the world's languages make systematic distinctions between nouns and verbs. Noun and verb categories have long been considered to be a universal feature of language cross-linguistically (Hockett, 1977; Hopper & Thompson, 1985), and have been posited to be a structural property found early in the evolution of language (Bickerton, 1990; Heine & Kuteva, 2007; Hurford, 2007). Research on developing communication systems such as homesign systems and emerging sign languages, suggest that early communicative systems show some distinctions between nominals and predicates, even when signal consistency across and within users is low (Goldin-Meadow et al., 1994; Goldin-Meadow et al., 2014). However, the universality of this distinction has been called into question, with arguments that some languages do not systematically distinguish between nouns and verbs (Evans, 2000; Gil, 1994). More recently, data from emerging sign languages suggest that consistent noun and verb categories, whether or not they are fundamental linguistic concepts, do not emerge fully formed (Abner et al., 2015, 2016; Tkachman & Sandler, 2013), but develop over time in a linguistic community.

In this chapter, I will investigate how categorical structures, such as nouns and verbs, emerge in innovated communication systems, and how such categories develop through use in interaction. Much of the current evidence for early noun-verb distinctions comes from emerging sign languages, which point to a gradual evolution of systematic category structures. I use an experimental investigation in the manual modality to understand how such distinctions might emerge in a new communication system, providing data comparable with that collected from natural languages. In the study described in this chapter, participants communicate, using only gesture, to describe a set of comparable scenes designed to elicit either verb-like or noun-like representations. Participants produce gestures both in isolation and during a communication task with a partner. I investigate how participants represent the objects presented in each scene type and whether they create systematic distinctions that resemble the nominal and predicate categories found in natural sign languages. Furthermore, I compare production in the individual with production in communication, to understand how pressures from interaction affect the forms participants produce.

3.1.1 Noun-verb distinctions in natural languages

Nouns and verbs can be classified both according to semantic similarities, or according to morphosyntactic features. Nouns generally reference objects, people and abstract concepts (though exceptions abound) and act as the agents and patients of a proposition. Nouns act as arguments of predicates, and often denote objects or concepts that act as subjects or objects in a clause. In spoken languages, nouns can be accompanied by determiners and demonstratives, and are frequently marked for gender, number or person. Verbs are classed semantically as referring to actions, states or processes, and tend to carry inflections or marking for tense, aspect and mood. In sign languages too, there exist particular formal differences between nominal and predicate signs that, though not universal, can be found cross-linguistically (Tkachman & Sandler, 2013). Verb signs are bigger than noun signs, using more proximal joints for articulation (Kimmelman, 2009), or may be longer in duration than nouns. The manner of movement (i.e. restrained/continuous) is used to differentiate signs of different word classes, verbs demonstrating continuous movement in comparison to nouns that demonstrate restrained movements (Supalla & Newport, 1978). Repetitions are frequently used to distinguish nouns and verbs, with nouns articulated with more repetitions. In cases where spoken languages can have an influence, mouthings borrowed from the community spoken language accompany nouns, distinguishing them as a sign class. Finally, handshape in sign languages may pattern differently in nouns than in verbs (Brentari, Coppola, Jung & Goldin-Meadow, 2013; Hunsicker & Goldin-Meadow, 2013). For example, ASL signers demonstrate consistent use of handshape for nominal signs, with specific handshapes used to consistently sign the same noun regardless of context. However, handshapes in classifier predicates vary depending on the context, such as whether an agent is present or absent (Brentari et al., 2013).

3.1.2 The evolution of category structure in emergent systems

It has been suggested that the classes of nouns and verbs emerge early in new linguistic systems (Goldin-Meadow et al., 1994; Goldin-Meadow et al., 2014) and that they may have been fundamental categories early on in the evolution of human language (Bickerton, 1990; Heine & Kuteva, 2007). Observation of emerging systems provides a view of how category structures emerge in novel linguistic systems, and how they evolve over time. In particular, the manual modality provides fertile ground to study the emergence of noun and verb categories. The unique social environments in which many deaf children and the resulting communities they form are situated allow the comparison of communication systems at different stages of evolution.

Homesign systems are gestural communication systems developed by deaf children, born to non-signing parents, who do not have access to a consistent linguistic model. Goldin-Meadow et al. (1994) analysed signs produced by a child homesigner, David, and found that he initially produced completely different signs for nouns and verbs. He later began to use similar forms that 'marked' a distinction in some way, such as through contrasting uni-directional and bi-directional movements, demonstrating some distinction between nouns and verbs.

Goldin-Meadow et al. (2014), studied nominal and predicate signs in Nicaraguan homesigners as well as signers from two consecutive cohorts of Nicaraguan Sign Language (NSL), an emerging sign language less than 50 years old. Similar to Brentari et al. (2013), they analysed handshape consistency in nominal and predicate signs and found differences between handshape patterns for nominals and predicates. The pattern, where handshapes for nominals tend to be consistently used for particular nouns, but handshapes for predicate constructions are more variable, appeared early in the language (Goldin-Meadow et al., 2014), though the patterns were not consistent across all users. Though the Nicaraguan homesigners in the study produced different signing patterns for nouns and verbs, consistency was variable across and within homesigners. From the first cohort of NSL, signers showed differential handshape patterning between nouns and verbs, demonstrating increased consistency both within and across-speakers, that increased again in the second cohort. The study also compared NSL signers with signs from ASL, which show robust differences in handshape patterns between nominals and predicates. NSL signing patterns resembled ASL signing patterns; both systems vary handshape contextually for predicates, but not nominals, though the distinction in NSL is not yet as consistent across signers as in ASL.

Additionally, Abner et al. (2015, 2016) assessed the strategies used to distinguish individual noun and verb signs in NSL, compared with Nicarguan homesigners. Using the video vignettes that were employed for the present study, homesigners and signers from three cohorts of NSL were asked to describe scenes in which a target object was used either typically or atypically, designed to elicit noun and verb signs for the same target element. For example, a typical use scene involving the object *camera*, showed a camera being used in a typical action, to take a photograph. An atypical use scene involving *camera* showed the camera object in an unusual context, such as being used to dig soil. The typical-use context, it is suggested, will lead to more verb-like signs; atypical-use scenes, highlighting the object in an unusual context, will elicit more noun-like signs for the object in the scene. All groups in the study showed some differences between their signs for typical scenes (target as noun) and atypical scenes (target as verb), marking a distinction between noun-like and verb-like expressions, but to different extents. NSL signers showed increasing contrasts between verb- and noun-like scenes over cohorts of the language, and delineated categories using strategies found across sign languages: range of movement and repetitions. By contrast, homesigners exibited differences in target sign order and the range of movements used in typical and atypical scenes, but did not show contrasts in repetition use to the same extent as later cohorts of NSL signers. The data collected from Nicaraguan homesigners and signers suggest that noun and verb categories can appear early in a communication system, but they become systematically distinct over cohorts of a language, highlighting the role of the linguistic community in the emergence of consistent sign categories.

Further evidence from emerging sign languages suggests that consistent noun and verb categories do not necessarily appear immediately in a language, but may develop through cultural evolution. Al-Sayyid Bedouin Sign Language (ABSL) is a sign language that emerged in a rural bedouin community in Israel during the 1930s, and does not appear to demonstrate systematic distinctions between nouns and verbs (Tkachman & Sandler, 2013). Though there are some tendencies to articulate verbs using more space than nouns, and to mark nouns with size-and-shape specifiers, neither strategy is used consistently to delineate a clear category structure. ABSL contrasts with Israeli Sign Language (ISL), of a similar age to ABSL, which does demonstrate robust distinctions between nouns and verbs. ABSL and ISL belong to two different classes of sign languages: village sign languages and community sign languages (Meir, Assaf et al., 2005). Village sign languages arise in small, closed rural communities with high rates of hereditary deafness. Community sign languages emerge when previously linguistically isolated deaf people come together, often due to changes in social policy such as the establishment of deaf schools or clubs. Deafness in community sign languages is less likely to be hereditary, or appear in multiple generations of a family; children are unlikely to be born into a family where other signers are present (Meir, Assaf et al., 2005). The social differences between village sign languages and community sign languages have been hypothesised to lead to different rates of linguistic evolution (Meir, Assaf et al., 2005; A. Senghas, 2005), serving to explain how two languages of the same age, and in the same geographical location, can exhibit category structure to such different extents. This difference in the socio-cultural context of the two languages has, in this case, led to strikingly different rates of systematisation, and further points to the role that cultural evolutionary processes play in the evolution of language. Even for categories such as nouns and verbs, that are considered fundamental in human communication, the communities in which those systems are used in interaction evidently shape the formal properties of each system.

The evidence from both NSL and ABSL indicates that categories for nouns and verbs may emerge early in a language, but that systematic contrasts take time to evolve, through use and conventionalisation in the community. In Nicaraguan homesign, the lack of a community does not allow the emergence of consistent, systematic noun-verb distinctions; the creation of the community that developed NSL led to the evolution, over cohorts, of consistent categories. Similarly, the contrast between ABSL and ISL suggests that, even when a linguistic community exists, consistent contrasts between nouns and verbs are not a given, but may be subject to the pressures of the socio-cultural context in which the language is used.

3.1.3 Investigating the emergence of noun and verb categories in the lab

The study detailed in this chapter aims to assess how category structures evolve in emerging manual systems, and to understand the role of production in the individual and in communication in the creation of categories proposed to be a fundamental feature of human linguistic systems. The experiment described in this chapter uses video vignettes previously used to elicit productions from Nicaraguan homesigners and signers of NSL by Abner et al. (2015, 2016). Participants were presented with scenes showing objects in different contexts: a typical-use scene in which the object is used in an expected action (such as *take photograph with camera*), and an atypical-use scene where the object is used in a non-typical action (such as *dig with camera*). Typical-use scenes were designed to elicit verb-like gestures, whilst atypical-use gestures, highlighting the object in an unusual context, were designed to elicit noun-like gestures for the target element.

The experiment uses the silent gesture framework as an analogue for manual communication, where hearing participants with no knowledge of sign language communicate the scenes described above, using only gesture and no speech. Silent gesture experiments provide two principal benefits: 1) by providing data in the manual modality to compare to natural sign language data and 2) by reducing the influence of participants' native language on the signals they produce. Silent gesture is a framework increasingly used to understand learners' preferences for linguistic constructions, without the influence of an existing language model (Goldin-Meadow et al., 2008; Hall et al., 2013; Schouwstra & de Swart, 2014). A principal aim of the study is to understand how signals for noun and verb expressions differ in individuals without communication and during communication. Participants in the experiment first produce gestures in isolation, before communicating with a partner about the target scenes. Participants completed a communication task (Fay et al., 2013, 2010; Garrod et al., 2007), taking turns to both produce and interpret gestures. Finally, participants once again produced gestures in isolation, to understand if changes to the system during communication persist after interaction is completed (Fehér, Wonnacott & Smith, 2016).

With the present experimental design, I will investigate the emergence of noun and verb categories both in the individual and in the community. Furthermore, using the silent gesture framework, I can directly compare data from experimental work with data collected from naturally emerging sign languages, further understanding the mechanisms that drive the emergence of linguistic structure. If nouns and verbs truly do represent "the most basic building blocks of language" (Bickerton, 1990), I predict that individual participants will represent typical-use and atypical-use scenes differently, showing a perceptual difference in the context of the target element. However, I do not expect that the strategies employed in sign languages to distinguish between nominal and verbal categories, such as range of movement and use of repetitions, will be used systematically at this stage. In constrast, during and following communication, I expect that participants will demonstrate greater contrasts between target elements in typical and atypical scenes, and further specify differences between target scenes to better aid communication. By contrasting production in the individual with production in a pair of communicators, I can test the immediate changes that communication has on sets of gestural signals, and whether such changes endure. I expect that gestural forms produced during and following communication will indicate the beginnings of a codified category structure that differentiates noun-like expressions from verb-like expressions, reflecting the distinctions found in natural languages.

3.2 Method

Participants took part in a task that first required them to produce gestures individually for a set of video scenes, before communicating about the same scenes with partner. Finally, we asked participants to produce gestures for the video scenes once again without a partner.

3.2.1 Participants

30 participants (11 male, 19 female, aged 18 to 27, median age of 20) were recruited from the University of Edinburgh's Careers Hub website. Participants were paid £7 take part in the experiment. All participants were self-reported to be right-handed native English speakers, with no knowledge of sign language. Participants took part in the experiment concurrently with another participant, who acted as their communication partner in stage 2 of the task (see section 3.2.3), giving a total of 15 pairs.

3.2.2 Materials

A set of 26 video scenes showing a number of target items were taken as a subset from a larger set of vignettes used by Abner et al. (2015, 2016). Target items in the scenes were either being used in a typical or atypical manner. For example, a scene in which a man takes a photo with a camera depicts a typical use of a camera, whereas a scene in which a man digs with a camera depicts an atypical use of a camera (shown in figure 3.1). Typical scenes were designed to elicit verb-like gestures, and atypical scenes were designed to elicit noun-like gestures. The set of scenes is organised such that each target item is depicted in a typical-use scene and at least one atypical-use scene. Each atypical-use is repeated with a minumum of two objects. Table 3.1 gives the full set of target scenes, shown as the pairing of objects with typical and atypical actions. Typical actions are actions the object is typically used for (e.g. taking a photograph with a camera, cutting with scissors, painting with a paintbrush). The video scenes were obtained from members of the Goldin-Meadow laboratory, at the University of Chicago, and were filmed in Chicago by members of the research group. For use in this experiment, the videos were converted from QuickTime files to mp4 files, to reduce file size and allow the video streaming software to stream videos without delay.

Participants were placed in individual experiment booths for the duration of the experiment. Video scenes and experiment instructions were presented on an Apple Thunderbolt monitor, attached to a Apple Macbook Air laptop. Recording was done via a Logitech webcam, also attached to the Macbook Air laptop. The experiment was

(a) Typical-use scene: take photograph with camera

(b) Atypical-use scene: dig with camera

Figure 3.1: Example video scenes for the target item *camera* used in the experiment. 3.1a shows a camera being used in a typical manner (i.e. taking a photograph); 3.1b shows a camera being used in an atypical manner (i.e. digging with a camera).

OBJECT	**ACTION**				
	Typical	**Atypical**			
		drop in bin	**drop in water**	**cover**	**dig with**
camera	take a photo	X			X
chair	sit			X	
door	open			X	
earring	wear		X		
hammer	hammer	X			
machete	cut with	X			
nail polish	paint nails		X		
paintbrush	paint wall	X			
ring	wear		X		
scissors	cut with		X		X
sewing machine	sew			X	
umbrella	open	X			

Table 3.1: Video scenes used as stimuli throughout the experiment. The video scene set comprises 12 objects, each used in a typical-use scenario. Each object also features in at least one scene of atypical use. There are four types of atypical use, each of which is used in at least 2 video scenes. There are 12 typical-use scenes (one for each object) and 14 atypical-use scenes, giving a total of 26 scenes.

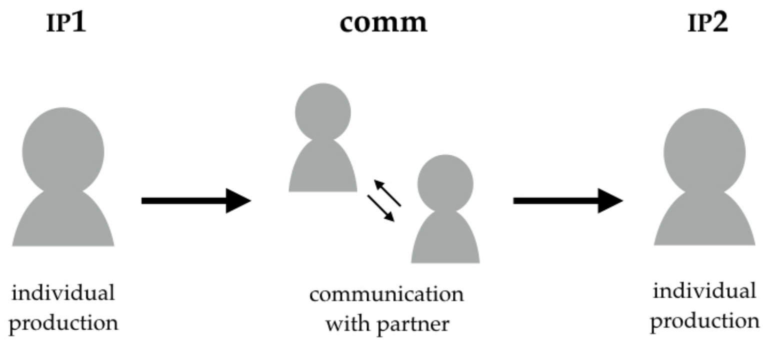


Figure 3.2: An illustration of stages in the experiment. Participants in IP1 and IP2 produce gestures to describe target scenes individually. In the middle stage, communication, participants communicate with a partner, taking turns to be **director** and **matcher**.

run using Psychopy (Peirce, 2007), and video streaming and recording between networked computers was enabled using VideoBox (Kirby, 2016), the custom software described in chapter 2.

3.2.3 Procedure

Participants took part in three stages. The first and third stages are termed **individual production stages**, where participants produced gestures for each video scene individually, without communicating with their partner. In the second stage, called the **communication stage**, participants communicated with their partner about the video scenes, through networked computers streaming the video feed from the webcams. Participants took part in the first individual production stage (IP1), before taking part in the communication stage (COMM) and then a second individual production stage (IP2), identical in procedure to the first, to assess the lasting effects of interaction on a set of communicative gestures. Participants were instructed not to use speech when gesturing, nor to use fingerspelling of any kind. Participants were also asked to remain seated throughout the task.

Individual production stages In the individual production stages (labelled here and henceforth as IP1 and IP2), participants were presented with each video scene in turn and were asked to produce a gesture to communicate what happened in the scene. At the beginning of each trial, participants were shown a 3 second countdown to the start of the trial, when they were presented with the video scene for that trial. The video scene was played twice in full, and then participants were presented with instructions to communicate the scene they had just watched to camera, using only gesture. Participants were again shown a 3 second countdown, to prepare them for recording. Whilst recording, participants saw themselves on screen in the VideoBox window, with their image mirrored. Text remained on screen throughout recording, instructing participants to press space bar when they had completed their gesture. Pressing the space bar stopped recording and completed the present trial and continued to the next trial. In each trial, participants were presented with, and had to communicate about, a single video scene. Participants completed trials for all video scenes in the set of 26, and the order of scenes was randomised for each participant. Individual production stages occurred at the beginning of the experimental session (IP1) and at the end of the experimental session (IP2), and both individual production stages followed the same procedure. The order of presentation of target scenes was randomised at each stage.

Communication stage In the communication stage (labelled as COMM), participants took turns with their partner to produce a gesture, as **director**, and to interpret their partner's gestures, as **matcher**.

As **director**, the participant was given a 3 second countdown at the start of the trial. They were then shown one of the video scenes from the set, twice through, as in the individual production stages. They were then instructed to communicate the scene to their partner, using only gesture, and given a 3 second countdown to prepare them for recording and streaming to their partner. During streaming, the director produced a gesture which was streamed to the networked computer operated by the matcher. As in the individual production stages, the director saw themselves mirrored onscreen, though the stream to the matcher was unmirrored. Either the director or the matcher could stop the recording and streaming by pressing space bar. When the stream was terminated, the director waited until the matcher made their guess at the correct interpretation of the gesture, after which both participants were given feedback (see below).

As **matcher**, the participant was shown a 3 second countdown at the beginning of the trial, followed by a short wait, in which they were presented with text reading *'Waiting for partner.'*, whilst the director was shown one of the video scenes. The matcher then received a 3 second countdown to prepare them for receiving the stream from their partner's camera. The matcher then saw their partner's gesture, unmirrored, onscreen and could terminate the recording and video stream by pressing space bar when they felt they understood the scene their partner was trying to communicate, and were ready to make their guess. When the stream was terminated, the matcher saw a set of 4 video scenes from which to make their interpretation, comprised of the target and three distractor scenes. The 4 scenes were selected as follows:

1. The target scene, with a **target object** and a **target action**.
   e.g. *take photo with camera*

2. A scene sharing the **target object** with the target scene, but showing a different action, the **foil action.**
   e.g. *drop camera in bin*

3. A scene sharing the **target action**[1] , but showing a different object, the **foil object**.
   e.g. *paint with paintbrush*

4. A scene depicting the **foil object** and the **foil action.**
   e.g. *drop paintbrush in bin*

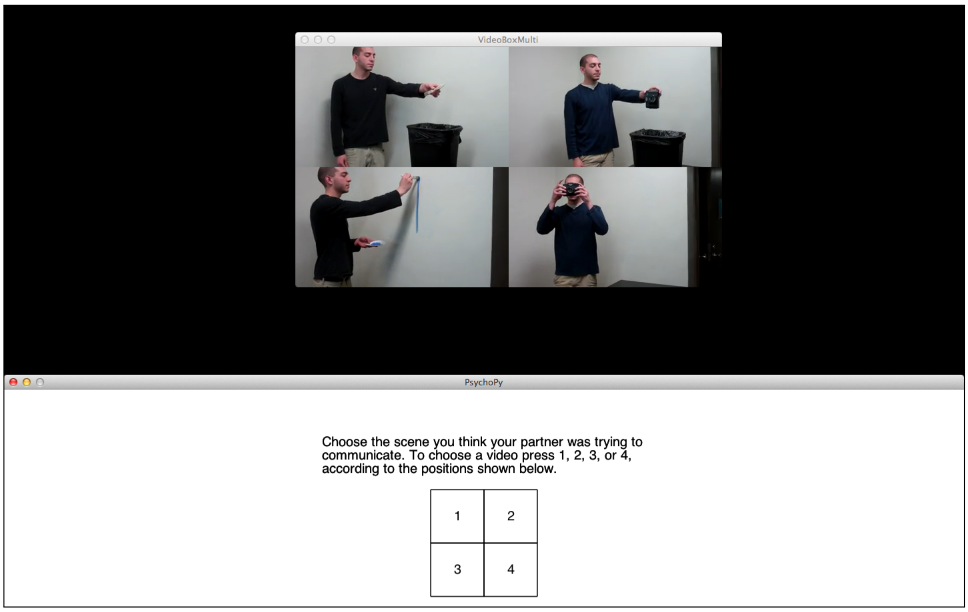


Figure 3.3: Screenshot of a matching screen. An array of four videos (target and three distractors) are shown in the VideoBox window (top), and instructions are given in the Psychopy window (bottom). Scenes depicted are *drop paintbrush in bin* (top left), *drop camera in bin* (top right), *paint with paintbrush* (bottom left), *take photo with camera* (bottom right).

[1]For typical actions, the comparable target action is another typical action. For example, *taking a photograph with a camera* is comparable with *cutting with scissors*

The target and distractor scenes were presented as a grid of 4 looping videos. The matcher had to press either the 1, 2, 3 or 4 key to make their choice depending on the position of the video, where 1 corresponds to the video in the upper left corner, 2 to the video in the upper right, 3 to the bottom left video and 4 to the bottom right video. A dummy grid showing which number to press for each position was presented on screen during interpretation. An example matching trial is shown in figure 3.3. The distractor videos looped until a guess was made. Once a guess had been made, both participants were given full feedback. For the matcher, if correct, the target video was highlighted in green. If the guess was incorrect, however, the target scene was highlighted in green and their selection was highlighted in red. For the director, if the matcher's guess was correct, the target video was shown on screen. If the matcher's guess was incorrect, both the target video and the selected video were shown on screen, with the target video highlighted in green, and the selected video highlighted in red. The director did not see the other videos in the distractor array presented to the matcher. Both participants were also presented with explicit confirmation of their success: text showing either "Correct" or "Incorrect" was presented in the Psychopy window for both participants. Feedback videos looped for 8 seconds before the experiment software continued to the next trial. The long feedback time was programmed to ensure that participants had time to watch both the target scene and the scene selected by the matcher. At the end of each trial, participants switched roles, so that the director became the matcher for the next trial, and vice versa. There were a total of 52 trials in the communication stage, so that each participant both directed and matched for all video scenes. Presentation order of the target scenes in each trial was randomised.

3.3 Results

3.3.1 Data coding scheme

Gesture videos were coded according to a number of features adapted from a coding scheme devised by researchers at the Goldin-Meadow laboratory to code signals produced by Nicaraguan homesigners and signers of NSL. The goal of the coding scheme is to capture differences between how participants gesture target items in typical scenes and target items in atypical scenes; as such, gesture coding focused on gestures for the target item in each scene. Coding classes can be grouped into a number of categories.

Coding for the number target signs that were expressed per trial and the position of the target sign in a gesture sequence assess how target items are broadly expressed and organised. Gestures were coded for number of target signs, which noted how many gestures depict the target element in the scene. For example, if a participant gestured a camera shape followed by a digging action for the scene *dig with camera*, then such a gesture would be coded as having 1 target sign, for the camera gesture. If, however, the participant only gestured the digging action, the gesture would be coded as having 0 target signs. For many of the typical scenes, the target item as object is indistinguishable from the target item as action (e.g. in figure 3.1, the camera object and the act of taking a photograph cannot be separated). In these cases, single gestures that combined expression of both object and action were coded as a single sign. Gestures were also coded for position of the target element in a gesture sequence for a target scene. Gestures could be either in **initial**, **medial**, or **final** position, or coded as **alone**, if gestures for the target element were the only gestures in the sequence.

Gestures were also coded for a number of formal features that contribute to linguistic structure in natural sign languages. These categories look more in detail at how participants gesture target items, noting the types of handshapes participants use, the use of a basehand to ground the target item gesture, and how participants use movement in their gestures.

Gestures were coded for handshape, according to the following categories:

- **Handling.** The handshape represents how someone would manipulate the object.
- **Instrument.** The handshape represents the object itself.
- **Descriptor.** The handshape represents features of the object, such as size or shape.
- **Deixis.** The participant indexes the object, usually using a gestural point (e.g. pointing at a finger on one hand to indicate a ring).
- **Object use.** The participants uses a real object (e.g. a ring, or the chair they are sat on) to stand in for the target object in the scene.
- **Lexical verb.** The handshape represents a lexical verb, such as 'sit'.

Gestures were coded appropriately for presence of a basehand. A gesture is considered to have a basehand if the gesture is two-handed, with a non-dominant hand providing grounding for the dominant hand. The non-dominant hand cannot depict the target element. For example, a gesture for *hammering on a wall* would have a basehand if the dominant hand represented the hammering action and a non-dominant hand represents either the wall, or the agent's hand resting on a wall. The basehand code is deemed relevant to the present study, as it provides grounding for a target-item gesture and provides an iconic representation of actions using particular targets (e.g. representing the wall, or the surface being hammered).

Gestures were also coded for movement features, such as the range of movement used and the number of repetitions of gestures for target elements. The range of movement was coded by joint, from the upper two finger joints up to full-body movement, identifying the size of the movement. After identifying the joints used in a movement, gestures were classified according to whether the movement corresponded to a path movement, which uses joints from the elbow to more proximal joints, or a local movement, using joints from the wrists to more distal joints. Path movements are larger movements in comparison to smaller, local movements. Finally, gestures were coded for repetitions, where a repetition is considered to be another iteration of any movement previously gestured in the same sequence (i.e. within a single production trial). The number of repetitions for each instance of the target element was noted.

Additionally, I investigated the interactive mechanisms that might lead to changes in how participants express differences between scene types in the experiment, by measuring participant alignment in each stage of the experiment. Alignment was measured in two ways: by measuring the difference in the number of target signs produced by each participant in a pair, and by measuring the difference between the types of gestures participants in a pair use (i.e. whether they use the same handshape to describe the same target item). These measures are detailed further in section 3.3.4.

I address the results of this chapter according to this coding scheme. Firstly, I discuss the broad features of the gesture sequences participants produce: the number of elements they produce in a sequence and the order in which they produce them. I then assess the formal properties of handshape type, basehand use, range of movement and repetitions to further understand whether participants create distinctions between scene types based on these criteria. Finally, I describe the findings that pertain to participant alignment, to further investigate the role that communication plays in shaping the gestures participants produce.

3.3.2 Expression of target elements

Number of target gestures. The number of gestures referring to the target element of a scene (as described in 3.3.1) were noted for each stage of the experiment. In IP1, individuals producing gestures in isolation frequently avoid explicitly gesturing the target sign in atypical scenes when it is possible to gesture other aspects of the scene.

(a)

(b)

Figure 3.4: Gesture sequences in which the target element is gestured (a) and where it is omitted (b). (a) depicts a gesture sequence for *take a photo with camera*, in which the participant produces a camera gesture. (b) depicts a gesture sequence for *drop camera in bin*, in which she simply gestures a dropping action, without specifying the object dropped.

For example, it is possible to depict the scene *drop camera in bin* without making the target object of the camera explicit, gesturing only a dropping action. However, gestures for *take a photo with camera* will typically use the camera object in the act of taking a photo; object and action are indistinguishable. Figure 3.4 illustrates the inclusion and omission of the target element with gestures for *take a photo with camera* and *drop camera in bin*.

When participants communicate in the second stage, they are more likely to explicitly gesture the target item, even in the atypical scenes where they omitted these gestures previously. In typical scenes, where participants nearly always produce at least one target sign, participants frequently include more than one target sign, further specifying the target scene (shown in figure 3.5). The task of communication leads to greater specificity and expressivity in describing the target scenes. Interestingly, this expressivity endures in the final stage, IP2. Although participants are once again gesturing in isolation, they do not omit target signs as they did for atypical scenes in IP1. Gestures remain expressive following communication and participants produce a similar frequency of target signs for target elements in each scene type.

Figure 3.6a illustrates the mean number of target signs for typical and atypical scenes at each stage. I ran a generalised mixed effects model with a poisson distri-

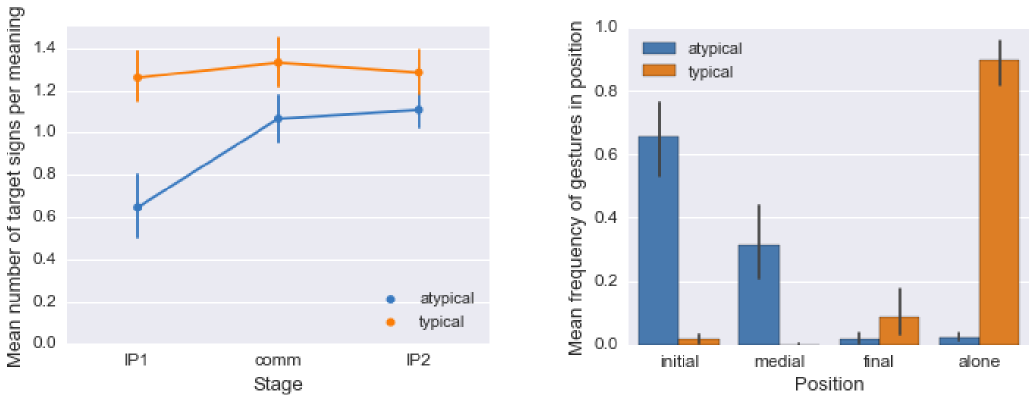
(a) (b)

Figure 3.5: A gesture sequence in which the participant produces two gestures for the target element. The gesture sequence depicts the target scene *using a sewing machine*. The participant produces a gesture of sewing with needle and thread in (a) and an additional gesture where her hands represent the sewing machine in (b).

bution for count data, using R (R Core Development Team, 2008) and lme4 (Bates et al., 2015). I included frequency of target signs as the outcome variable, with stage, scene type and their interaction included as fixed effects. Participant, pair and target scene were included as random effects with random slopes, and the random effects structure for participant was nested within pair[2]. The model revealed a significantly better fit over the reduced model ($\chi^2 = 23.85, p < 0.001$). Analysis of the model revealed an effect of stage ($\beta = 0.25, SE = 0.04, z = 6.81, p < 0.001$) and an effect of scene type ($\beta = 0.58, SE = 0.10, z = 5.67, p < 0.001$), as well as a significant interaction between stage and scene type ($\beta = -0.24, SE = 0.05, z = -4.89, p < 0.001$). Participants frequently omit gestures specifying the target element in IP1, but do not omit target elements during and after communication with a partner. Target elements are omitted from productions for atypical scenes more frequently than typical scenes, where object and action are gestured simultaneously, and gestures for typical scenes generally show higher frequencies of target signs than gestures for atypical scenes.

Target element position Figure 3.6b shows the proportion of gestures in each position in a gesture sequence. Target elements in atypical scenes were most frequently gestured in initial and medial positions, and gestured without other elements (i.e., alone) in typical scenes. Using a binomial generalised mixed effects model, I analysed the effect of scene type and stage on the presence of target element signs gestured alone (those that were not part of a larger gesture sequence). The model revealed an improved fit over the reduced model that used only scene type as a fixed effect ($\chi^2 = 22.44, p < 0.001$), and a model including the interaction between scene type and stage showed no improvement over the model without

[2]The random effects structure described here is used throughout this chapter, unless otherwise specified. Models were run using bound optimisation by quadratic approximation (given with parameter name *bobyqa*).



(a) (b)

Figure 3.6: Target sign frequency (a) and target sign position (b). (a) shows the mean frequency of target element gestures at each stage, for atypical and typical scenes. Participants omit target elements in atypical scenes in IP1, but not in typical scenes. During and following communication, participants no longer omit target item gestures, regardless of scene type. (b) illustrates the mean proportion of gestures in each position (initial, medial, final, alone) in a gesture sequence for a target scene. Target elements in atypical scenes are gestured in initial and medial positions, whereas target elements in typical scenes are overwhelmingly gestured alone, without other elements. Error bars represent 95% confidence intervals, sampled by pairs in the experiment, here and throughout this chapter.

the interaction term ($\chi^2 = 3.40, p = 0.07$). The model demonstrated a significant effect of scene type ($\beta = 8.47, SE = 0.68, z = 12.49, p < 0.001$), and of stage ($\beta = -0.69, SE = 0.15, z = -4.57, p < 0.001$). The frequency of target signs gestured alone decreases over stages, as participants' gesture sequences become more expressive, including more gestures per trial. Participants also show a significant preference for gesturing target elements alone when they are gesturing typical scenes than when they gesture atypical scenes, suggesting that, for typical scenes, the target object and the target action are gestured simultaneously, or possibly treated as one and the same. For target elements in atypical scenes, the object is treated as object-like, and is differentiated from the action it is subject to. This leads to strings where the target element (here as an object) is gestured, followed by the action such as *drop* or *dig with*, producing gesture sequences of the form (S)OV.

Handshape representation Analysis of handshapes used revealed that participants overwhelmingly preferred handling handshapes over other types of handshapes, as shown in figure 3.7. Participants used handling shapes to represent target elements in 74.8% of all trials.

However, handling handshapes were not the only type used, all gesturers pro-

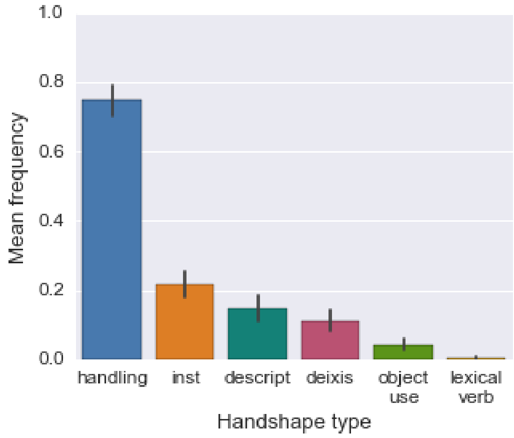
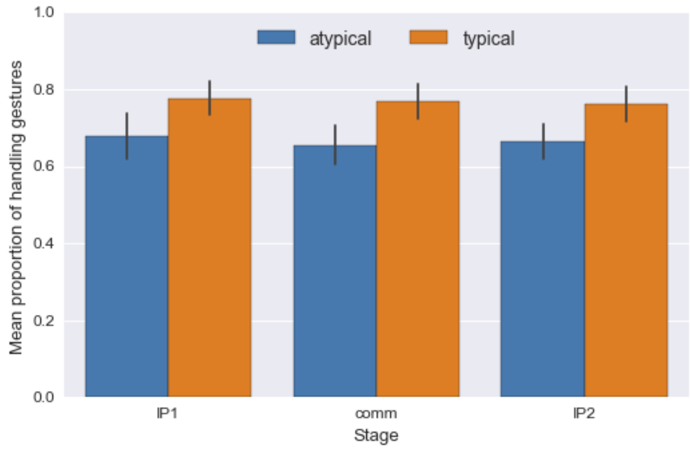


Figure 3.7: Mean frequency of different gesture types described in 3.3.1. Participants overwhelmingly prefer handling gestures to other gestures, but do use a range of different gesture types.

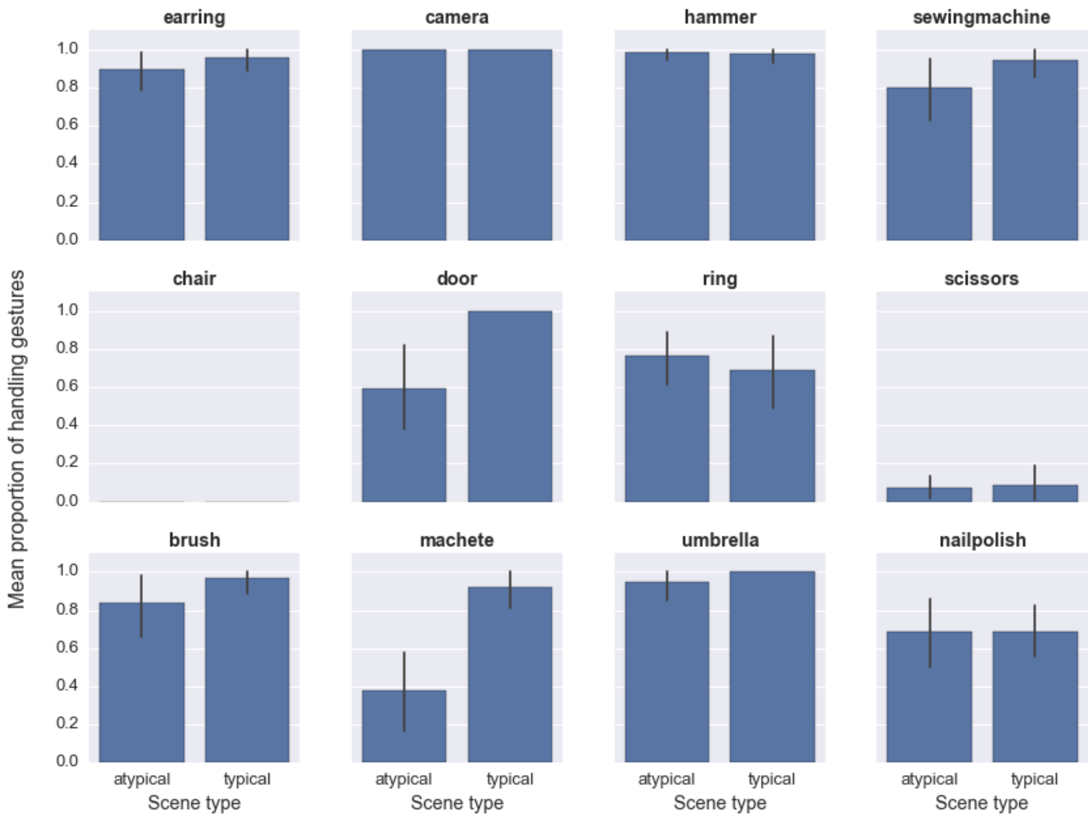
ducing other handshape types at least once in production, in addition to handling gestures, to represent the target element. These results are consistent with previous studies suggesting a preference for handling handshapes amongst hearing gesturers (Padden et al., 2014).

Figure 3.8a, shows the mean proportion of handling gestures per meaning at each stage of the experiment, and for each scene type. Further inspection of the data (illustrated in figure 3.8b) indicates that use of handling handshapes is largely driven by the target elements. Participants use handling handshapes consistently for each target item, in both typical and atypical cases, but exhibit variation across target items.

A generalised binomial mixed effects analysis assessed the effect of scene type on the use of handling handshapes. The model demonstrated no improvement over the null model ($\chi^2 = 0.89, p = 0.35$). Participants show a preference for handling handshapes across all scene types, with no significant distinction made between typical and atypical scenes. However, when participants do make a distinction based on scene type, it is in the expected direction, with handling gestures used more in the typical cases than the atypical cases. Similar handshapes are re-used across scene types for the same target element, rather than systematically distinguishing between scene types. For typical scenes, this generally entails representing the whole scene iconically. For example, figure 3.9a (top) shows a participant's gesture for *paint nails with nail polish*, where she iconically represents the target scene. In her gesture for *drop nail polish in water* (3.9a(bottom)), she re-uses the handling gesture, painting her nails, before gesturing a dropping motion. The nail-painting does not actually occur in the target scene; although still an iconic gesture, it is not here an iconic represent-



(a)



(b)

Figure 3.8: Proportion of handling gestures per meaning for each scene type across each stage of the experiment (a), and for each scene type for each target element in the experiment (b). Where (a) suggests a greater preference for handling gestures in typical scenes than atypical scenes, (b) indicates that use of handling gestures varies more between target elements than between scene types.

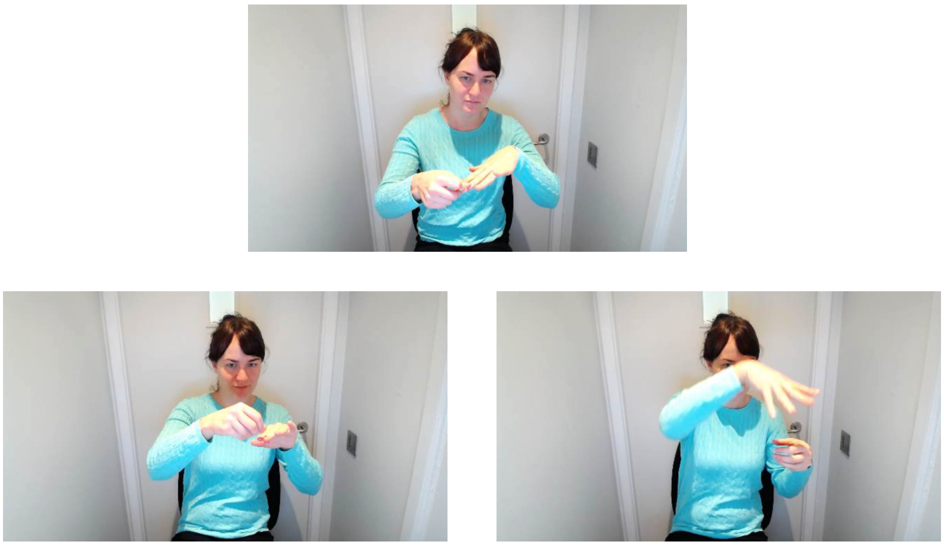
ation of the target scene. Participants however, may use different handshapes when representing the whole scene iconically. For example, figure 3.9b depicts two gestures for the typical and atypical uses of machete. The participant's gesture for *cut with machete* (top) iconically represents the action depicted in the scene, in which a man holds a branch and cuts it using a machete. In the corresponding gesture for the atypical scene, *drop machete in bin* (bottom), the participant does not re-use the handling gesture but uses a descriptor gesture that outlines the size and shape of the machete. The gesture iconically represents the machete, and acts more as an iconic description of object in the target scene, in which a man holds the long machete from the handle and drops it into a large bucket. These results suggest that the types of gestures used are not used to systematically differentiate between the different contexts in which objects are used, but are driven by iconic representation, even when participants are communicating.

Basehand gestures Gestures including a basehand, as described in section 3.3.1, were present more frequently in gestures for typical-use scenes than in gestures for atypical-use scenes (see figure 3.10b).

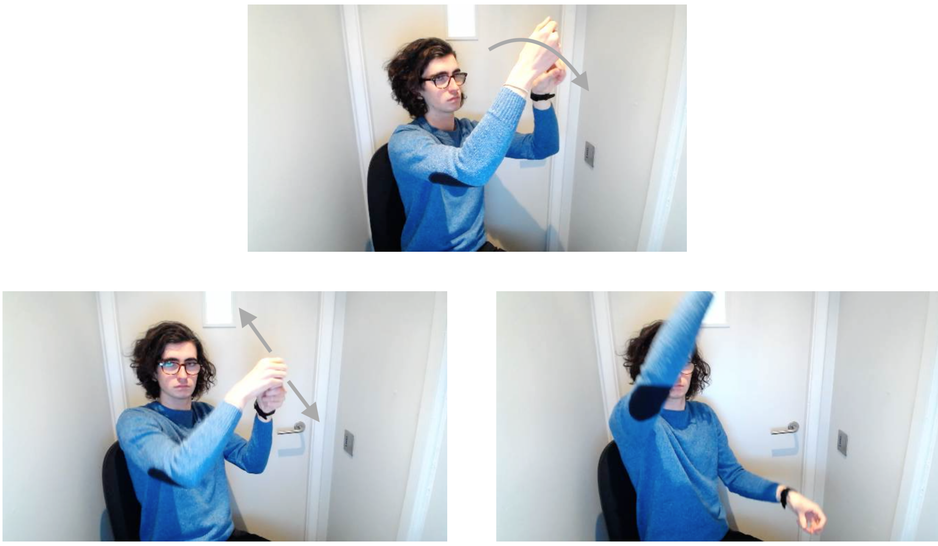
I analysed the effect of scene type and stage, as well as their interaction, on the presence of basehand gestures, using a binomial generalised mixed effects model. The model including the interaction term showed a significant improvement over the reduced model ($\chi^2 = 4.13, p = 0.04$). Analysis of the model demonstrated an effect of scene type ($\beta = 3.35, SE = 1.26, z = 2.65, p = 0.008$), and an effect of stage ($\beta = 0.65, SE = 0.11, z = 6.20, p < 0.001$), as well as a significant interaction between scene type and stage ($\beta = -0.34, SE = 0.16, z = -2.07, p = 0.04$). Use of basehand gestures increases over stages in the experiment, and participants use basehand gestures significantly more often for typical-use scenes than for atypical-use scenes, though this difference reduces to a small extent over stages of the experiment. Basehand gestures provide an iconic grounding for a subset of actions in the target scenes, and therefore help to iconically represent the scene as whole (for example, when representing the wall in gestures for *hammering a wall*, as in figure 3.10a).

3.3.3 Range of movement and repetitions

Range of movement Gestures were coded for joint use, where more distal joints (in the wrist and hands) are involved in small, local movements and proximal joints such as the elbow and shoulders are coded as producing large, path movements. Figure 3.11 shows the proportion of path and local movements for each scene type at each stage of the experiment.

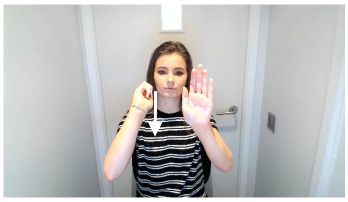


(a) Gestures for *paint nails with nail polish* (top) and *drop nail polish in water* (bottom). Both gestures use a handling gesture that suggests someone painting their nails. The gesture for *drop nail polish in water* adds a dropping gesture to show dropping the nail polish into a glass of water. The nail painting gesture represents the object, while the drop gesture represents the action. The participant does not represent the object and action simultaneously.

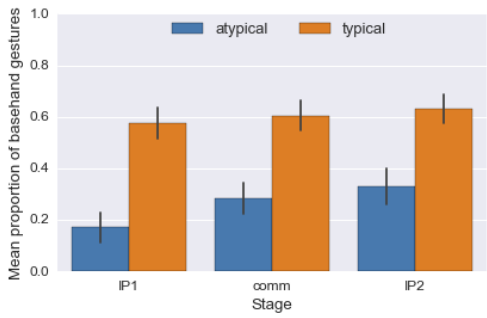


(b) Gestures for *cut with machete* (top) and *drop machete in bin* (bottom). For *cut with machete*, the participant uses a handling gesture to represent the cutting action. For *drop machete in bin*, the participant gestures the shape and length of the machete, using a descriptor gesture.

Figure 3.9: Examples of the different handshape strategies used to describe different scenes. In (a), the participant uses the same handling gesture across both meanings, but adds a further gesture in the atypical instance to communicate the action in the scene. In (b), the participant uses different types of gestures to represent the typical (top) and atypical (bottom) use of the object, a machete. All examples here are from the communication stage.

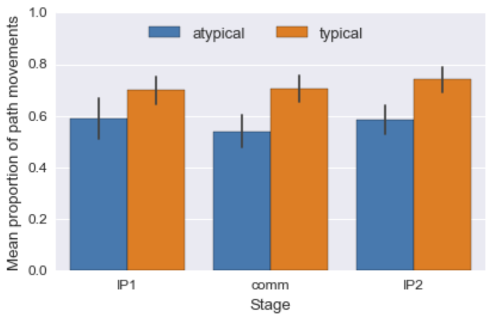


(a)

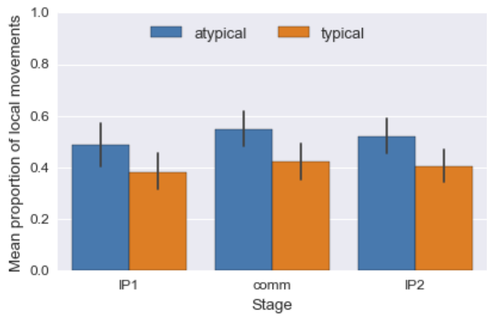


(b)

Figure 3.10: (a) Example of basehand gesture for *hammering on a wall*. The flat palm represents the basehand, the fisted hand represents the hammer object. (b) Mean proportion of basehand gestures over stages in the experiment, for each scene type. Use of basehand gestures increases over stages, and basehand gestures are used to communicate typical scenes more frequently than atypical scenes.



(a)



(b)

Figure 3.11: Proportion of path movements (a) and local movements (b) per meaning at each stage of the the experiment, for each scene type. (a) indicates a greater preference for path gestures in typical scenes than atypical scenes. In (b), the opposite trend is suggested; participants show a greater preference for local gestures in atypical scenes than typical scenes.

A binomial generalised mixed effects analysis investigated the effect of scene type and stage on the proportion of path gestures. The model showed a marginal improvement over a reduced model including only scene type ($\chi^2 = 3.93, p = 0.05$). Analysis of the model revealed no effect of scene type ($\beta = 2.71, SE = 1.55, z = 1.74, p = 0.08$), and a significant effect of stage ($\beta = 0.22, SE = 0.11, z = 2.05, p = 0.04$), suggesting an overall increase in the use of path gestures over stages of the experiment. Figure 3.12 illustrates the proportion of path and local gestures for each target element in the experiment. As with handshape, participants use movement to iconically represent target elements rather than distinguish between the use of target elements systematically. Large objects, such as machete (figure 3.9b), tend to be gestured with path movements, compared to smaller objects like nail polish (figure 3.9a), which tend to be gestured with local movements. Movement size is not used consistently to distinguish between target elements in different scene types.

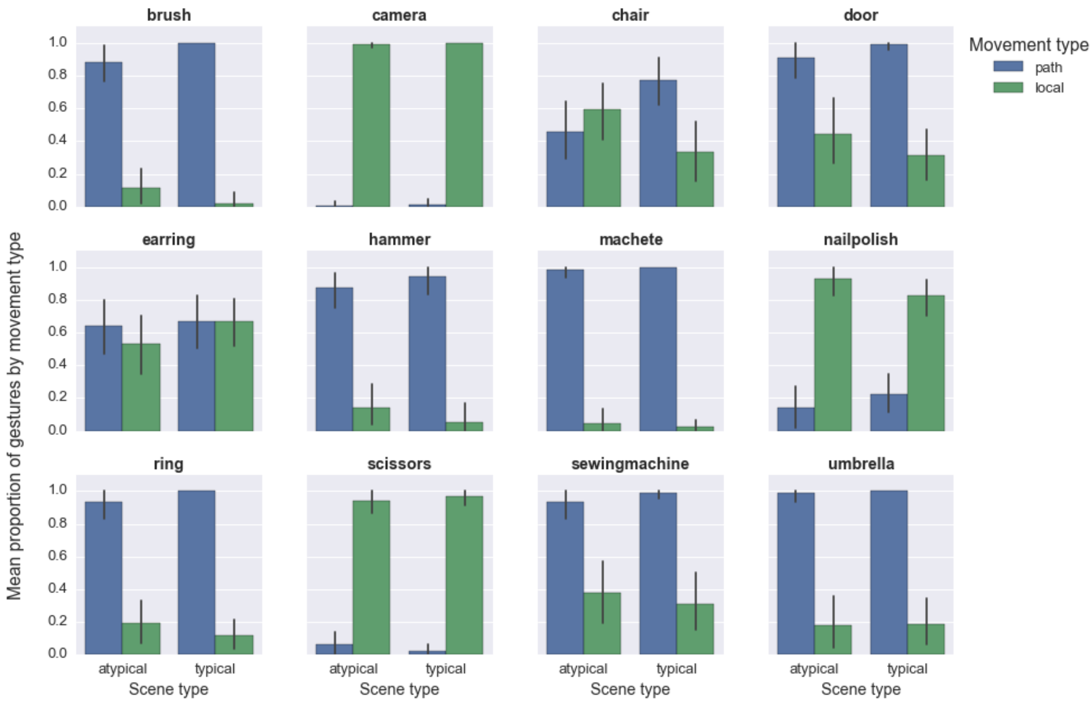


Figure 3.12: Mean proportion of path (blue) and local (green) movements per meaning for each target element in the experiment, for each scene type. There is no systematic difference in the use of path or local movements between scene types. Instead, participants use different movement types to iconically represent target elements (e.g. larger objects are gestured with path movements, smaller objects are more likely to be gestured with local movements.

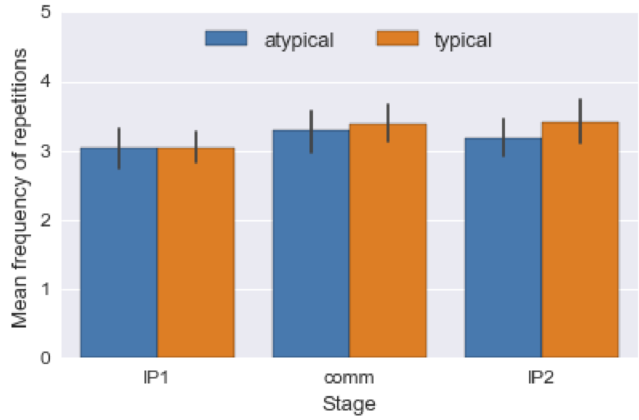
Repetitions Participants commonly used repetitions in sequences of gestures for a particular target meaning. The frequency of repetitions in gestures for each target meaning was noted and analysed at each stage in the experiment (see figure 3.13a). A generalised mixed effects analysis using a poisson distribution investigated the effect of stage and scene type on the frequency of repetitions within a gesture string. The model revealed no improvement over the reduced model ($\chi^2 = 0.004, p = 0.95$) and a model including only stage as a fixed effect also did not show improved fit over the null model ($\chi^2 = 3.47, p = 0.62$).

Participants show no systematic use of repetitions to differentiate between scene type. Variation in the use of repetitions is driven largely by the target element (figure 3.13b), suggesting that participants use repetitions to iconically represent elements of the scene. For example, gesture sequences representing nail polish (for targets of *paint nails with nail polish* and *drop nail polish in water*, see figure 3.9a) frequently involve repetition, as the target scene itself depicts a woman painting her nails, an action necessarily repeated in order to paint all the nails on her hand. Similarly, the action of cutting a sheet of paper with scissors, as depicted in the target scene *cut with scissors*, requires a repeated cutting action. Importantly, such iconic representations do not account for the full range of use of repetitions. Participants frequently repeat full sequences of gestures, or repeat a target element gesture that does not have repeated movement in the target scene. For example, participants frequently gesture putting on a ring to communicate target scenes that depict both *put on ring* and *drop ring in water*, and frequently repeat this action (i.e. they put on the ring, and then take it off and put it back on again), even though it is not repeated in the target scene (see figure 3.14b). In addition to iconic representation, repetitions are used to emphasise target elements, though this is not used to systematically distinguish between different scenes.

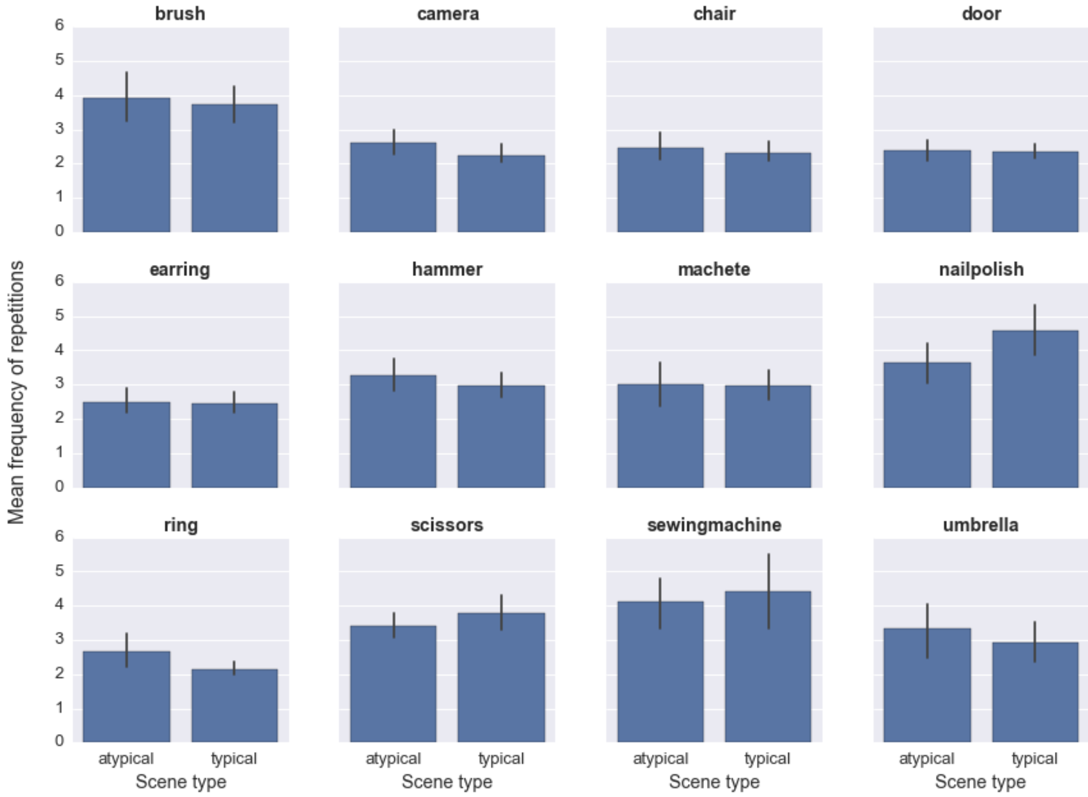
3.3.4 Alignment

Participants' gestures show greater expressivity from the communication stage of the experiment, producing more target elements per meaning and producing fewer gestures for target elements in isolated sequences.

I measured alignment between pairs of participants to understand whether pairs negotiate a shared system, even as those systems include more gestural elements. Alignment was measured in two ways: by calculating the difference between the number of gestures for target elements participants produced, for each pair, and by calculating the Levenshtein distance between the codes given for sets of target elements participants produced in their pairs (for example, whether both participants produce handling gestures, or instrument gestures). One potential issue in comparing the number of target signs each participant produces is that, if the number of



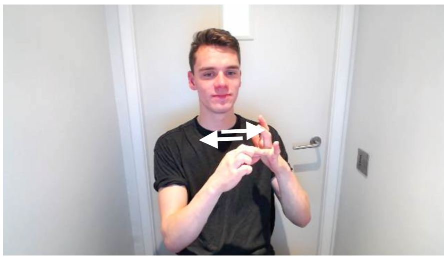
(a)



(b)

Figure 3.13: Mean frequency of repetitions per meaning for each scene type across each stage of the experiment (a), and for each scene type for each target element in the experiment (b). Participants show differences in the use of repetitions across target items but not between scene types, suggesting that repetitions are used to iconically represent items in the target scenes, rather than as a category marker.

(a) Gesture for target scene *cut with scissors*



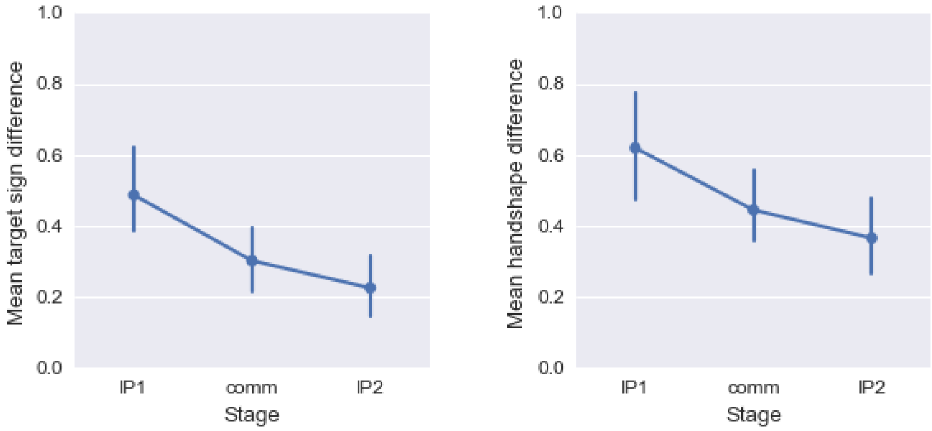
(b) Gesture for target scene *put on ring*

Figure 3.14: Gestures for *cut with scissors* (a) and *put on ring* (b), showing use of repetition. In (a), the participant re-creates the repetition inherent in the target scene, of a repeated cutting action. In (b), the participant repeats the gesture of putting on the ring, though this repetition is not found in the target scene.

target elements simply reduces over the time, the smaller number of elements will lead to a smaller difference between participants, whether or not they are aligned. I point to two findings that suggest this is not the case: that, on average, the number of target signs increases during and following communication, and that the types of gestures participants produce become more similar, as described below.

Target element difference I calculated the difference between the number of target elements each participant produced, in their pairs. For example, if one participant produced 3 target elements for a particular scene, and their partner produced 1, their difference score is 2. Figure 3.15a shows the mean target difference at each stage of the experiment. A generalised mixed effects model using a poisson distribution for categorical data investigated the effect of stage on the difference between number of target elements produced, including pair and target scene as random effects with random intercepts. The model demonstrated an improved fit over the null model ($\chi^2 = 39.98, p < 0.001$), and revealed a significant effect of stage ($\beta = -0.40, SE = 0.06, z = -6.31, p < 0.001$), indicating a decrease in the difference between the number of elements communicating partners produced after IP1. Participants demonstrate increasingly similar frequencies in the number of target signs produced, showing convergence on the length of gestures for scenes in the experiment.

Difference of gesture type The difference between the types of gestures communicating partners produced was calculated as the Levenshtein distance between sets of gestures for each target scene. The Levenshtein distance is an edit-distance measure that notes the minimum number of insertions, deletions or substitutions that would transform one string into another. The Levenshtein distance was calculated on strings



(a)

(b)

Figure 3.15: (a) Mean difference in the number of target elements produced in pairs, and (b) mean Levenshtein distance between the sets of gestures pairs of participants produce. Participants produce increasingly similar gestures for target elements, both in the number of elements produced and the types of gestures they use to represent target elements.

of codes for the target elements for a target scene, assessing the similarity of those codes within a pair. Figure 3.15b shows the mean Levenshtein distance at each stage of the experiment. I used a generalised mixed effects model with a poisson distribution to analyse the Levenshtein distance at each stage of the experiment, including pair and target scene as random effects with random intercepts. The model revealed a significant improvement over the null model ($\chi^2 = 15.11, p < 0.001$) and exhibited a significant effect of stage ($\chi^2 = 0.24, SE = 0.06, z = -3.94, p < 0.001$). The types of gestures participants produce become increasingly similar across stages of the experiment. The decrease in the difference between the number of target elements communicating partners produce, as well as the reduction in the difference between gestures produced for target elements suggests that participants become increasingly aligned within their pairs after communication, despite the increase in the number of gestures produced per target scene.

3.4 Discussion

The present study sought to understand the effect of communication on the distinctions participants make when gesturing a set of objects in different contexts. I presented participants with objects in typical-use contexts and atypical-use contexts, aiming to elicit gestures that would represent verb-like expressions and noun-like expressions, respectively. Participants took part in two stages in which they produced gestures in isolation, and an intermediary stage in which they communicated with a partner. I compared the gestures that participants produce in communication with a partner to those that participants produce in isolation.

The results described in this chapter suggest that communication shapes the gestures that participants produce. Participants' gestures become more expressive during communication, and indeed, participants produce the highest frequency of target element gestures in communication, though expressivity is preserved in the second creation stage. Communication leads to more expressive signals that serve the communicative act (Carr et al., 2016; Culbertson & Kirby, 2016; Kemp & Regier, 2012; Kirby et al., 2015; Regier et al., 2015), further specifying the target element in this case to ensure successful communication. The effects of interaction persist into the second creation stage, following communication, indicating that interaction between interlocutors can have a lasting effect on the system (Fehér et al., 2016), with changes that occur in communication being preserved in production contexts that do not include an interlocutor. The gestures participants produce show further adaptation to communication, as participants become more aligned with their partner after communication, both participants showing convergence on similar systems.

Though these findings demonstrate some effects of communication, they do not demonstrate the emergence of systematic category structures from communication alone. Participants show variation in their gestures along a number of criteria found in natural sign languages, such as handshape and range of movement, but these criteria are used to iconically represent items in the target scenes, regardless of context. Participants do not show systematic variation of these strategies dependent on the context in which the target item appears (i.e. in typical-use or atypical-use contexts).

For instance, participants show a preference for handling gestures over other gesture types, a finding that replicates previous silent gesture study (Padden et al., 2014), demonstrating an overall preference for handling gestures. Participants do not consistently use handshape to distinguish items in different target scenes. However, for target items that do indicate a consistent difference in the use of handshape (*door* and *machete* in 3.8b), the distinction is in the expected direction, with typical-use scenes producing more handling handshapes than atypical-use scenes, producing a verb-like gesture that represents the action of using the target item (Hunsicker & Goldin-Meadow, 2013; Padden et al., 2013). This distinction is not made systematically across target elements, and participants frequently use iconic representations of target elements, re-using the same strategies for each object across scene types. The use of consistent strategies by target rather than by scene type occurs for handshape as well as the range of movement and the number of repetitions used. For example, bigger movements are used to gesture larger objects, and repetitions are more likely to be used in scenes where an action is necessarily repeated, such as scenes for *nail-polish* or *scissors*.

However, participants do employ some strategies to distinguish between scene types in the experiment. Participants order target elements differently, producing target elements in typical scenes in isolation, whereas target elements in atypical scenes are often part of a larger sequence of gestures, and usually occupy an initial or medial position in the sequence. In typical scenes, action and object are gestured simultaneously; in atypical scenes, the target element is abstracted and gestured as an object, followed by the action in the scenes, such as *drop* or *dig*. The ordering of elements in this way is consistent across the experiment, and suggests firstly that participants view the target elements in each scene type differently, but may also suggest that consistent ordering of strings of gestures to depict a scene can emerge earlier in a communication system than the systematic differentiation of formal properties of gestures.

Secondly, for target scenes where use of a basehand to ground the target element is possible, such as those involving *hammer* or *brush*, basehand gestures appear more frequently to communicate typical scenes than to communicate atypical scenes. Use of the basehand further indicates that action and object are gestured simultaneously, with the basehand being a part of that iconic representation. In atypical scenes, where basehand use is less frequent, the target element is abstracted as an object. For example, for the atypical-use target scene *drop hammer in water*, the hammer is frequently gestured with a hammering action, although there is no hammering action in the target scene. However, omission of a basehand gesture frequently distinguishes this scene from the typical-use scene of *hammering on a wall*; the gesture no longer communicates the act of hammering, which involves gesturing the surface the agent is hammering, but the object, *hammer*. However, it is important to note that basehand gestures are not likely to be gestured for all target elements, only for those where grounding, such as a wall or other surface, is involved in the action. Therefore, omission and inclusion of basehand gestures does not mark distinctions between scene types across the system of gestures, but across a subclass of target items. The difference in the use of basehand gestures serves to highlight the perceived difference between target elements in the two contrasting scene types.

Previous research supports the present results, suggesting that communication alone is not sufficient for the emergence of systematic structure, producing a system that marks differences and similarities consistently across a range of meanings. Rather, linguistic structure is a trade-off between pressures for expressivity and learnability that are applied by communication and transmission to new learners, respectively (Kirby et al., 2015; Motamedi et al., under revision). Communication leads to informative systems that allow interlocutors to communicate accurately, and lead to communicators converging on shared, increasingly symbolic systems that are both informative and communicatively efficient. Although communication alone can lead to regularisation and convergence on consistent word orders (Schouwstra et al., 2016), as the present results suggest, it does not lead to the systematic use of structured category markers. The results presented here are also consistent with comparable studies investigating signs produced to describe the same stimuli in Nicaraguan homesigners and signers of NSL. Abner et al. (2015, 2016) found the emergence of noun and verb categories over cohorts in NSL signers, the categorical structures becoming more contrastive and consistent with each cohort. Contrastive noun and verb classes in this case take repeated use in communication and transmission to new learners of the language to stabilise and become entrenched. Nicaraguan homesigners, similar to participants in the present study, show a word order preference for targets in typical and atypical scenes, with atypical targets less likely to be gestured in final position compared to typical targets. Though the homesigners demonstrate some contrast in the use of movement size, showing larger movements for targets in typical-use scenes, they also demonstrate patterns suggesting by-target representation rather than consistent distinctions patterned on scene type. They demonstrate some contrast in the frequency of repetitions in each scene type, but only for target scenes that include repetitions as part of the event. However, Abner et al. (2015, 2016) do find a contrast in the range of movement in homesigners, whilst there is no contrastive use of movement in my participants. Homesigners, unlike participants here, use their signs as their primary communication system. Though family members of homesigners do not tend to use homesigns as their primary communication system, and may not even use the system fluently, homesigners do use their signs to communicate with family members. As such, homesign systems may represent a stage between signals produced by individuals in isolation, and signals produced between equal communicators.

Finally, these findings contrast with previous results produced in the graphical medium (Fay et al., 2013, 2010; Garrod et al., 2007), as the gestures participants produce become more expressive, but not symbolic or communicatively efficient. Symbolic communication requires common ground between participants, shared reference allowing iconic forms to become more symbolic. In the present study, participants only communicate for one round, with participants building common ground within communication, but not to the extent that gestures become notably less iconic.

It is possible that further rounds of communication would allow further development of the signals to symbols. However, the iconicity of a manual-visual medium, particularly one that privileges the gesturer's body as a representational device, cannot be ignored, and may offer an explanation as to the retention of iconicity in participants' gestures. Communication, by itself, leads to more expressive gestures, that allow communicating participants to be to be mutually understood. Participants in interaction show the beginnings of convergence on a shared system, demonstrating alignment on the length of target element gestures and on the handshapes they use to describe target elements; however, such minimal interaction does not afford them the opportunity to build the common ground required to develop symbolic, communicatively efficient signals.

Participants in isolation produce short gestures that frequently omit specific aspects of the target scene. During and following communication, participants produce gestures that are more expressive and demonstrate alignment with their partners. Communication does not lead to systematic distinctions between the contexts in which target elements are used, though the seeds of such distinctions are present in the ordering of gestures for target elements, suggesting that participants account for the differences between scene types. The results presented in this chapter support and expand on data collected from an emerging sign language, Nicaraguan Sign Language, suggesting that individuals are sensitive to fundamental differences between noun-like meanings and verb-like meanings, but that these differences do not immediately become codified in a language. Instances of communication set the wheels in motion for language-like signals, but do not provide the pressures necessary to develop gestures that are communicatively efficient or systematic.

3.5 Summary

The motivation for the study detailed in this chapter comes from data collected from emerging sign languages, investigating the structures that appear in a languages earliest stages of evolution. However, natural language data is currently unable to provide a clear view of the mechanisms that lead to the emergence of these structures, leaving many open questions about how cultural processes, such as communication, shape language. The methodology presented here offers a way to tease apart the effects of solo production and communication, providing data that is directly comparable with data collected on naturally emerging gesture systems, and bridging the gap between laboratory and field data.

I have contrasted silent gesturers producing gestures in isolation, without communication, with silent gesturers using gesture to communicate with a partner. Participants gesturing in isolation produce short gestures which are easy to produce, but become more expressive during and following communication. Participants show differences early on in the experiment in how they order gestures in typical-use scenes (verb-like) and atypical-use scenes (noun-like). However, the gestures participants produce exhibit few systematic formal differences between the target elements in typical-use scenes and in atypical-use scenes. Instead, the gestures participants produce rely heavily on iconic representation throughout the experiment. Although minimal communication leads to more expressive gestures, previous work suggests that repeated communication is required to drive the emergence of symbolic systems, and that communication alone is not sufficient to drive the emergence of systematic category structure.

Chapter 4

The cultural evolution of complex constructions: denoting who does what to whom

4.1 Introduction

4.1.1 The problem of sign language verbal agreement

A large proportion of the world's languages demonstrate linguistic relationships, by way of agreement, between a predicate and its arguments. Verb agreement signals a relationship between the predicate and the semantic roles of the participants in a clause, denoting who does what to whom. It is a feature prevalent cross-linguistically, with approximately 80% of documented spoken languages showing verbal person marking on either the agent role, the patient role, or both (Siewierska, 2013). Sign languages, too, demonstrate systematic linguistic tools to denote the semantic and syntactic roles between person and predicates. Cross-linguistically, sign languages employ spatial contrasts, or spatial modulation, to note the relationships between participants in a sentence (Mathur & Rathmann, 2012; Sandler & Lillo-Martin, 2006), which is frequently signalled by indexing referents in the space around the signer. For example, in the English sentence given in (1), the verb *ask* agrees with the third-person singular subject (in the agent role) using the inflectional form *-s*. For a similar construction in British Sign Language (BSL), shown in (2), signers use referential indices to refer to participant roles. The signer points to one location in the signing space to represent the subject, and to another to represent the object, and the verbal form, ASK, moves between these locations to signify who is asking whom.

1. **English**

He asks her.
he ask-3SG me

2. **BSL**
HE$_a$ $_a$ASK$_b$ HER$_b$

Though there are differences between sign languages with respect to which verbs allow indexing of this kind, and how indexing is implemented, the use of spatial modulation to denote arguments in discourse is attested across most sign languages (Mathur & Rathmann, 2012). As visuo-spatial languages, it is unsurprising that sign languages make use of the visual space around the signer, across a number of linguistic levels (Padden, 1990), and it has been suggested that the iconic possibilities of gestural spatial mappings on to real world events make the use of space almost inevitable in signed languages (Aronoff et al., 2005). Indeed, the beginnings of systematic directionality in predicate signs have been found not only in young sign languages such as Nicaraguan Sign Language (NSL; A. Senghas & Coppola, 2001), and Al-Sayyid Bedouin Sign Language (ABSL; Padden et al., 2010) but also in homesign systems (Goldin-Meadow, 2003).

However, though these systems are present across sign languages and the rudiments of spatial grammars show up in early sign systems, the evolution and development of systematic spatial modulation is not necessarily straight-forward. Spatial agreement systems encode complex relationships between the arguments of a clause and necessitate the reanalysis of iconic space as grammatical. Spatial agreement forms develop late in comparison to other morphologically-rich constructions in spoken languages, such as Turkish. The age of acquisition for American Sign Language acquirers occurs between 3;0 and 3;6, comparable to acquirers of English, a language without rich verbal agreement (Pichler, 2012). For constructions that denote non-present referents, development can continue even later, and are not fully mastered by age 5 (G. Morgan, Herman & Woll, 2002). Furthermore, although the beginnings of spatial modulation are evident in some early sign systems, there is evidence to suggest that the use of systematic contrastive space does not emerge immediately in a language, but takes time to develop. Research into two emerging sign languages, Israeli Sign Language (ISL) and Al-Sayyid Bedouin Sign Language (ABSL), indicated that neither sign language has a system of spatial modulation in early generations of signers (Meir, 2012; Meir et al., 2007; Padden et al., 2010). However, these systems are changing, to the extent that younger signers (i.e. of more recent generations) show an increase in the consistency of spatial contrasts to denote who does what to whom (Padden et al., 2010), though neither language demonstrates systematic use of space to extent that older sign languages (such as ASL and BSL) do. Padden et al. (2010) suggest that the slow development of consistent spatial grammar may be the result of competing iconicities. The iconicity inherent in the signer's body, that can be exploited to represent animate agents (particularly human agents) sits in opposition to the iconicity that allows participants to occupy contrastive locations. Body iconicity, grounded in the signer's anatomy, appears early. The evolution of a spatial system that can contrast two animate participants in different locations removed from the body requires the abstraction of the grammatical concept of *person*, and thus may take longer to emerge. According to this account, a grammatical system where any animate agent can occupy subject position conflicts with one where subject is grounded in the signer's own body (Meir et al., 2007; Padden et al., 2010).

As a system that relies on a visual representation of space, we can also ask how spatial modulation develops. Would such as system rely heavily on iconic representation before being abstracted, or do some semantic constructions necessitate agreement forms? Verbal agreement systems in sign languages do not effect the whole class of verbs, but are applicable only to a subset of verbs. Padden (1990) identified three types of verbs that operate in sign languages: plain verbs, that do not inflect for agreement, spatial verbs that represent a locative relationship, and agreement verbs that inflect, or denote person through referent indexing. Typical agreement verbs can involve physical transfer between participants in a sentence, such as GIVE, but also frequently involve non-physical transfer, as is the case with ASK. Spatial verbs such as MOVE and CARRY-BY-HAND can also use space to denote the movement of participants in space, but do signal a spatial relationship in the real world. The data collected from ISL and ABSL signers suggests that there is no immediate distinction in how different verb types are signed; all types rely on iconic mappings between participants in early generations of both languages. Therefore, it is possible that development over time leads to divergence between verb types. Spatial verbs, that are more likely to present an iconic mapping onto real space (e.g. *Hannah runs to Sarah*), may rely more heavily on iconic mappings that reflect the particular discourse context. Agreement verbs, on the other hand, frequently denote events of transfer that do not necessarily signal a real spatial mapping between agent and patient (e.g. in *Hannah gives a book to Sarah*, Sarah does not have to be present for Hannah to give her the book). Abstraction from iconic representation to more arbitrary spatial reference is required (Padden et al., 2010), and it is this abstraction that takes time to emerge.

Finally, it is certainly not uncontroversial that the use of space in sign languages is fully grammatical and fully encoded in the linguistic system. Liddell (2003) argues that the spatial grammar of ASL is not a fully linguistic phenomenon, but combines morphological units (verb forms) with non-linguistic gestures (index points). In particular, he argues that, because points can index any part of the signing space, the range of possible loci are uncountable, and therefore undefinable as a set of morphemes. Similarly, Lillo-Martin and Meier (2011) posit a system that conjoins linguistic morphemes with gestural points, though in a different way, proposing that spatial reference in ASL corresponds to the combination of spoken pronouns and gestural points in spoken language users, such that, in sign languages, the combination of linguistic reference (pronominal index) and gestural reference (locative point) are simultaneously articulated as a single deictic point. Furthermore, data collected from BSL signers by Cormier, Fenlon and Schembri (2015) indicate that spatial mappings even in agreement systems may not be as arbitrary or as systematic as previously thought. BSL signers demonstrated high levels of iconic mappings between referents in agreement verb constructions, as well as high use of body iconicity (as attested in ABSL and ISL) and relatively low levels of arbitrary spatial mappings. Cormier et al. (2015) (among others, including Liddell (2003)), reject the term agreement, carrying as it does the assumption that such a system would be a) fully linguistically encoded and b) formally similar to agreement systems in spoken languages. Agreement verbs are otherwise called indicating verbs (Cormier et al., 2015; Liddell, 2003), or directional verbs (Lillo-Martin & Meier, 2011). These arguments, however, are beyond the scope of this chapter, and for the sake of consistency, I will refer to these systems and the verbs they affect in terms of agreement, though the concerns detailed here will be discussed further in section 4.4, in light of the experimental results.

The present state of the literature on spatial modulation in sign languages suggests that this phenomenon poses two principal problems to linguistic enquiry; the mechanisms by which such a system develops, and how a mapping with high iconic potential integrates into an encoded linguistic system. Here, I reformulate these problems into two main research questions.

1. How do complex morphological constructions emerge and evolve in a language?

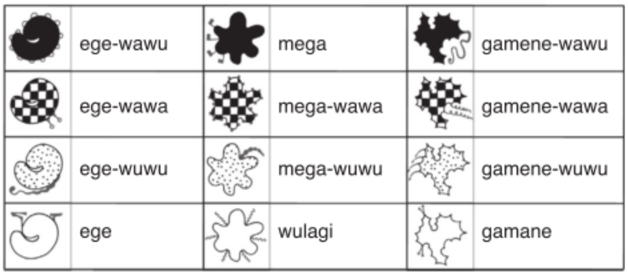
2. How does the iconicity afforded by a visual medium interact with the development of systematic structure in a linguistic system?

In this rest of this chapter, I will motivate and detail an experiment that poses an investigation into these two research questions. Participants communicate with a partner using only gesture, to describe a set of events that involve the interaction of two animate participants, *Hannah* and *Sarah*. Successful communication relies on distinguishing between and signalling the relationship between the two participants, or denoting who does what to whom.

4.1.2 Experimentally investigating the evolution of complex linguistic structure

Previous experimental research has highlighted the importance of cultural evolutionary processes in the emergence of linguistic structure (Carr et al., 2016; Fay & Ellison, 2013; Fay et al., 2010; Kirby et al., 2015; Silvey et al., 2014; Theisen-White et al., 2011). Interaction between users of a system and transmission to new learners of a system shape language to reflect the demands of the task (see chapter 2) and the biases of the learner (Culbertson & Kirby, 2016; Perfors & Navarro, 2014; Thompson, Kirby & Smith, 2016). For the most part, experiments employing cultural evolutionary frameworks have focussed on simple concepts and meanings, such as objects that share physical properties (Carr et al., 2016; Kirby et al., 2015) or concepts that share certain semantic properties on one or two dimensions (Fay et al., 2010; Theisen-White et al., 2011). The signals that evolve to communicate these meanings, therefore, reflect this relative simplicity. Though they exhibit increasingly compositional or combinatorial structure, the structures themselves remain relatively simple, sequentially concatenating small numbers of elements (usually two or three) that reflect the dimensions of the meanings space (see examples in figure 4.1 below). Even with a complex meaning space, such as the essentially infinite triangle meaning space used by Carr et al. (2016), participants reduce the possible dimensions they can signal to a small set; in the case of Carr et al. (2016), participants signal the most salient features of the triangles, size and shape.

The results of this body of work show the emergence of structured communication from unstructured holistic signals, used to communicate simple meanings. However, spatial reference in sign languages is used to describe the relationship between animate participants in complex events, showing complexity that has not previously been investigated experimentally. The slow rate of evolution of spatial structures in emerging sign languages (Padden et al., 2010), and possibly in older sign languages (Cormier et al., 2015), as well as the relatively late acquisition of these structures by children (Pichler, 2012), underline the complexity of spatial reference in manual communication systems. Much of the recent study using silent gesture, where hearing participants must communicate using only gesture and no speech, does focus on event signalling rather than item signalling, to assess how participants order event information in a signal (Goldin-Meadow et al., 2008; Hall et al., 2013; Schouwstra, 2016; Schouwstra & de Swart, 2014), providing a useful tool to study the biases of



(a) Example language reproduced from Kirby et al. (2015).

(b) Example signal from chapter 2, for the meaning *prison*.

Figure 4.1: Signals from cultural evolution experiments that reflect the lack of complexity in target meanings.The example in (a) taken from Kirby et al. (2015), shows a set of signals that systematically distinguish target meanings, on the two dimensions of shape and pattern. (b), from the experiments described in chapter 2 shows a two-part signal for *prison* that reflects the two-dimensions of the target meaning (theme: law, function: location).

language users that may underlie the typological distributions of different word orders cross-linguistically. However, the majority of silent gesture research thus far has focused on production in the individual at a given time, and not on the cultural evolution of gestural systems. The methodologies used in these studies remain unsuitable for investigating a construction as complex as the spatial reference systems found in sign languages. Word order does not demonstrate the same level of complexity as spatial agreement. Possible word orders are much more constrained than spatial reference systems, and learners show clear preferences for particular word orders, namely SOV and SVO (Goldin-Meadow et al., 2008; Schouwstra & de Swart, 2014). Word order has been shown to emerge early in young sign languages (Sandler et al., 2005; A. Senghas et al., 1997) and is suggested to be consistent even in child homesigners (Goldin-Meadow, 2003). In contrast, evidence from emerging sign language indicates that cultural evolution is instrumental in both the initial emergence of a spatial reference system as well as its ongoing systematisation. A basic word order for ABSL developed by only the second generation (Sandler et al., 2005), whilst consistent use of spatial reference was not present even after three generations (Padden et al., 2010). With these considerations in mind, it is evident that any experimental model of spatial reference should take cultural evolutionary processes into account.

Spatial modulation in sign languages is a device considered to rely on iconic spatial mappings between clause arguments. Cultural evolutionary models of language evolution have also investigated the presence and effects of iconicity and systematicity in signals, as well as how the two mechanisms interact as languages evolve. Theisen et al. (2010) and Theisen-White et al. (2011) used a graphical communication task to assess this interaction. Participants were asked to communicate a set of meanings through drawings, which allowed iconic representations of concepts. Their findings suggested that as the sets of signals participants produce develop systematic structure, sharing parts of the signals that reflect similarities in meanings, there is a reduction in iconicity. The implication, that languages are more likely to have high levels of iconicity in earlier generations, before systematicity develops, is corroborated by data from sign languages, which suggest that older sign languages demonstrate less iconicity than emerging sign languages, and show a reduction in iconicity over time (Aronoff et al., 2005; Frishberg, 1975; Padden et al., 2010). Furthermore, experimental findings indicate that the presence of iconicity can hamper the evolution of systematicity in the early stages of a communication system (G. Roberts et al., 2015; Verhoef et al., 2015). G. Roberts et al. (2015) manipulated the ability to produce iconic signals between pairs of communicating participants, and found that allowing iconic mappings led to signals that were complex and lacked combinatoriality, compared with the development of combinatorial signals in a condition that did not allow iconic signals. Verhoef et al. (2015) extended this design to a transmission framework, using slide whistles as the medium of articulation. Participants in transmission chains learnt whistles produced by a previous participant, either in a condition where iconic mappings between signal and meanings was retained, or in a condition where signals were scrambled, interfering with any iconic mappings between signal and meaning. Though both contexts eventually led to signals exhibiting combinatorial structure, systems in which signals retained iconicity showed slower rates of evolution than signals in the non-iconic condition. Sign languages utilise the iconic potential of a manual-visual system to represent relationships between arguments in a clause. Such high levels of iconicity early on in a communication system may hinder the evolution of systematicity in the use of space, and may serve to explain the shape of spatial reference systems cross-linguistically.

The experiment I detail in this chapter aims to investigate how spatial reference systems evolve in the early stages of a communication system; in particular, I attempt to understand the roles of interaction and transmission in combination in the evolution of systematic spatial reference. Furthermore, the prevalent iconicity in spatial reference system lends itself well to experimental study of how iconic signals change over time, and how those signals change as sets of gestures become increasingly systematic. I implement an experimental design that investigates the two principal research questions formulated above: how do these constructions emerge and evolve as gestures are used and transmitted to new learners, and how does the potential for iconic signals shape the gestures participants produce? Participants in the experiment communicate with a partner about a set of events, that include one or both of two event participants, *Hannah* and *Sarah*. Participants are shown pairs of events, presented as orthographic sentences in English, and asked to communicate them using only gesture. The sentence pairs are arranged into two different types: one in which the agent across both sentences is the same (same-agent) and one in which two agents in a pair of sentences are different (different-agent). Participants must create a system that allows them to correctly identify the arguments in a sentence. The gestures participants use to communicate become the training for another pair of participants in a transmission chain structure, modelling the transmission of communicative signals from one generation to another. Finally, within the possible events participants are exposed to, I also varied the types of verbs participants had to communicate, reflecting the possible verb types classified in sign languages. Sentences could describe a verb event of motion with no end point, a verb of motion with a specified end point or goal, a physical transfer event or a non-physical transfer event. With this experimental design, I can examine the cultural evolutionary mechanisms that allow the evolution of complex linguistic constructions out of non-linguistic communicative signals. I provide an experimental account of the evolution of spatial reference that is informed by data on natural sign languages, and which attempts to shed light on the obscurities of a complex and highly iconic system.

4.2 Methods

Participants were recruited for an experimental study in which they were presented with target sentences in English and communicated them using only gesture. Participants were trained on gestures produced by a previous participant before communicating with a partner in the testing stage.

4.2.1 Participants

50 participants (15 male, 35 female, aged 18 to 42, with a median age of 22) were recruited from the University of Edinburgh's Careers Hub website. Participants were paid £7 to take part in the experiment, which took between 30 and 50 minutes to complete. All participants were self-reported right-handed native English speakers with no knowledge of sign language. Participants who had taken part in previous similar experiments (e.g. a silent gesture task) were not allowed to participate. Participants were organised into five diffusion chains of five generations, with a pair of participants at each generation (see figure 4.3).

4.2.2 Materials

Throughout the experiment, participants were asked to communicate pairs of sentences, presented orthographically in English. In training and matching trials, participants watched a previously recorded video of another person gesturing and had to choose out of an array the pair of sentences they thought the person in the video was trying to communicate. Sentences in the experiment involved two actors, Hannah and Sarah, who could either be the agent of the sentence, be the goal or end location of the sentence, or not be present at all in the sentence. Verbs in the sentence were from one of four categories, shown in table 4.1; verbs in the experiment were all presented in the present progressive form.

Pairs of sentences only used verbs from the same category, and sentence pairs were grouped into sets of four, so that each set contained one pair of each verb type. Two types of sets were used: same-agent sets where the agent of each sentence in

Verb type	Description	Verbs
plain spatial verbs	Verbs of motion with no specified end point	to cycle to run to swim to walk
spatial locative verbs	Verbs of motion with specified end point, P	to walk to P to run to P to swim to P to walk to P
physical transfer verbs	Verbs denoting a physical transfer of an object to a recipient, R	to kick a ball to R to give a book to R to send a letter to R to throw a hat to R
non-physical transfer verbs	Verbs denoting a non-physical or metaphorical transfer to goal or recipient, R.	to help R to phone R to praise R to scold R

Table 4.1: List of verbs used in stimulus sentences in the experiment. Verbs are grouped according to verb type, and throughout the experiment verbs were presented in the present progressive form (e.g. *Hannah is kicking a ball to Sarah*).

the pair was the same agent (e.g. Hannah was the agent of both sentences), and different-agent sets where the agent of each sentence was different (e.g. Hannah was the agent in the first sentence, Sarah the agent in the second). Figure 4.2 illustrates an example of the sentences used in the experiment and how they are structured in pairs and sets (a list of all sets used in the experiment is given in appendix A). Spatial reference in sign languages often occurs across a stretch of discourse, such that indexed referents can be understood as attached to a referential location and referred to over a stretch of discourse (Lillo-Martin & Meier, 2011). As such, the pairing of sentences in each trial of the experiments represents a stretch of 'discourse', or communication, in which the same arguments (*Hannah* and *Sarah*) may appear. Blocks of four allow individual participants, as directors, the possibility of using gestures across trials, representing larger stretches of discourse in which interlocutors communicate about the same arguments.

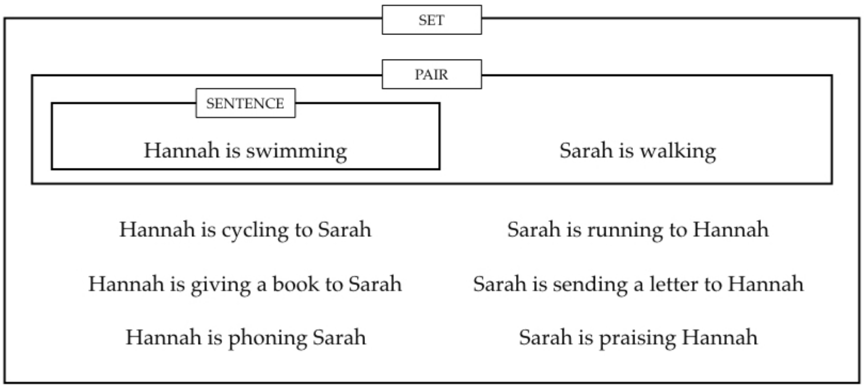


Figure 4.2: Examples of sentences used in the experiment, and how they are structured. At each trial, a pair of sentences is presented. Each pair uses two verbs of the same category, as given in table 4.1. Pairs of sentences are grouped into sets of four, where one pair exhibiting each verb type is present. The full body of sentences comprises four sets, 2 same-agent sets and 2 different-agent sets (given in appendix A). The current example is a different-agent set.

For matching trials in training and testing, participants were presented with an array of four sentence pairs to choose from. One pair was the target pair, and the other three foil pairs either differed from the target on the verbs used or the agent configuration (either different-agent or same-agent), or on both. I provide an example below, where, for the target pair given, the three foils would include those given in the following list.

Target pair:

Hannah is helping Sarah.
Sarah is scolding Hannah.

Foils include:

1. One pair that retained the same agent order as the target, but used different verbs from the same category.

   Hannah is praising Sarah.
   Sarah is phoning Hannah.

2. One pair that used the same verbs but only used one name in agent position.

   Hannah is helping Sarah.
   Hannah is scolding Sarah.

3. One pair that differed from the target on both the verbs and on the agent configuration.

   Hannah is praising Sarah.
   Hannah is phoning Sarah.

Foil verbs comprised the two remaining verbs in the same category, and were randomly placed in either the first or second sentence in a pair. Foil agent selection differs depending on the target agents. For same-agent targets, the other agent that is not used in the target will be used as the agent in one sentence of the foil pair, but the order of the sentences in the pair is randomised. For different-agent target pairs, one of the agents (either Hannah or Sarah) will be randomly selected to appear as the agent twice in the foil pair. For all sentences, agent and goal are never given the same name (i.e. *Hannah is praising Hannah* never appears).

Participants were placed in individual experiment booths throughout the experiment, in front of an Apple Thunderbolt monitor with an affixed Logitech webcam. Monitor and webcam were connected to an Apple Macbook Air laptop running Psychopy (Peirce, 2007) and VideoBox (Kirby, 2016), the custom software described in chapter 2, allowing streaming and recording via networked computers.

4.2.3 Procedure

Participants were organised into five chains of five generations, similar to the design used in experiment 1 described in chapter 2 (see figure 4.3). Participants from generations two to five took part in both a training and a testing stage, and participants in generation one, as the starting participants in each chain, only took part in the testing stage. This differs from the experiments described in chapter 2, where seed gestures were used as starting systems for each chain. Here, generation 1 participants must innovate the system, allowing the investigation of changes in a system from innovation in the individual, to the use and transmission of that system.

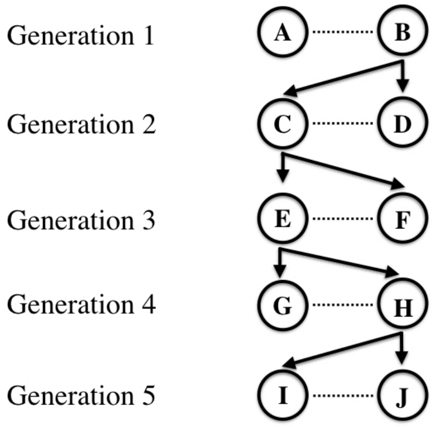


Figure 4.3: Transmission chain structure in the experiment. Pairs of participants communicate with each other at each generation. Participants in generation 1 innovate gestures in communication. Participants in generations 2 to 5 are trained on gestures produced by one of the two participants in the previous generation, before communicating with their partner. Solid lines represent transmission, dashed lines represent interaction.

Training stage

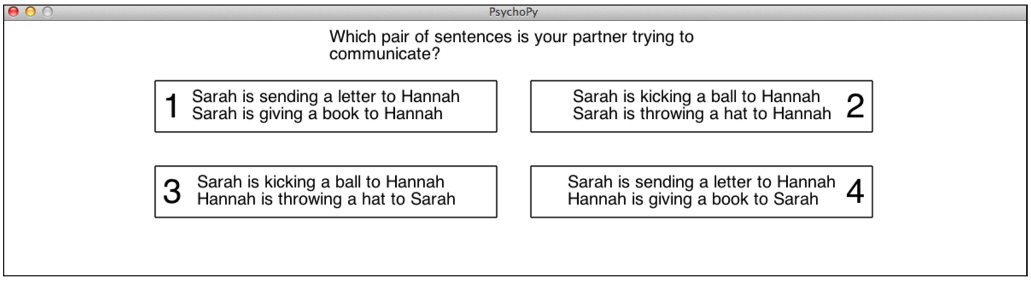
Participants in the training stage were trained on gestures produced by a model participant in the previous generation of the same chain. The model was randomly selected from one of the two participants in the previous pair, and all gesture videos produced by the model participant were shown over trials in the training stage. Participants were given a three second countdown to prepare them for the start of a trial, at which point the training video would be shown on screen, with the matching array presented in a Psychopy window beneath the video stream. The matching array was presented with each pair shown in a separate box, each box labelled with a number from 1 to 4 (see figure 4.4a), designating its position on screen. The position of each pair was randomised at each trial. Participants could make a guess at any point whilst watching the training video by pressing the number key corresponding to a box containing a pair of sentences. If participants made a guess before the end of the video, the video stopped streaming, showing a black VideoBox window; VideoBox also presented a black window if the video finished playing before the participant had made their guess. Upon making a guess, participants were given feedback, and presented with text on-screen noting whether their guess was correct or incorrect. The target pair was highlighted with a green box, and, if incorrect, the participant's selection was highlighted with a red box. Feedback was presented for five seconds. Participants took part in a total of 16 training trials, one for each target pair. Participants took part in the training stage individually, and the order of presentation for sentence pairs was randomised for each participant. Participants in the first generation of each chain did not take part in the training stage, but only completed the testing portion of the experiment.

Testing stage

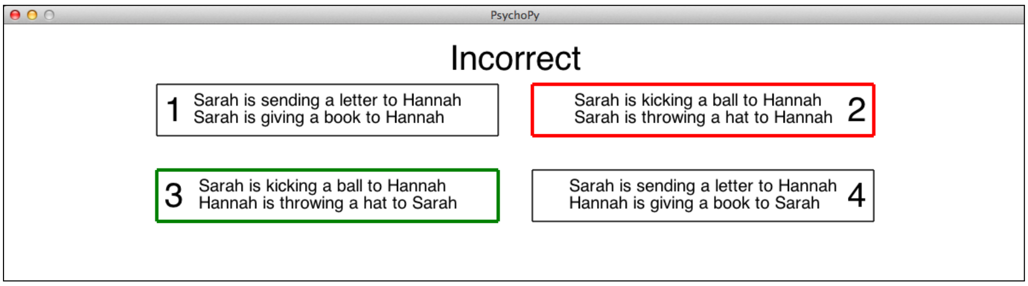
During the testing stage, participants remained in individual experiment booths and interaction was facilitated through video streaming between networked computers. Participants in the testing stage communicated with a partner using only gesture, taking turns to produce and interpret gestures in a director-matcher task (Garrod et al., 2007; Healey et al., 2007). Participants both produced and interpreted gestures for all 16 sentence pairs in the experiment, giving a total of 32 testing trials. Participants held the same role, as either director or matcher, for four trials at a time, completing a full set of sentence pairs, before switching roles with their partner. Each set of sentences was presented as a block in the experiment, and participants were notified of their role at the beginning of each block and at the beginning of each trial. Each participant held the role of director for all sets of target sentence pairs, and the order for each director was randomised, as was the order of sentence pairs within a set.

In the role of **director**, participants were presented with a pair of sentences and instructed to communicate both sentences to their partner, using only gesture. They were presented with a three second countdown at the beginning of the trial, and then shown the pair of sentences in the Psychopy window onscreen for five seconds. They were then presented with another three second countdown to prepare them for recording and streaming to their partner. When the recording and streaming started, participants saw themselves in the VideoBox window, with their image mirrored, though the stream to their partner remained unmirrored. The target sentence pair was again shown onscreen during recording and streaming. The director could stop the recording and streaming by pressing space bar, upon which video streaming was terminated for both director and matcher. The director then waited for the matcher to make a guess, before receiving feedback on the success of the trial. They were shown whether or not the matcher's guess was correct or incorrect, and presented with both the target pair (highlighted in green) and the matcher's selection (highlighted in red if incorrect).

In the role of **matcher**, the participant had to interpret gestures produced by their partner. The matcher was given a three second countdown at the start of the trial and then shown text reading "Waiting for partner" whilst the director was shown the target sentences. A three second countdown prepared the participant for streaming



(a)



(b)

Figure 4.4: Screenshots showing presentation of a matching array during a testing trial, before a selection has been made (a), and after a selection has been made (b). Pairs of sentences are labelled with numbers 1 to 4, and participants press the corresponding number key to make a choice. Upon selection, the target meaning is highlighted in green and the matcher's selection, if incorrect, is highlighted in red. Text positioned above the sentence pairs gives explicit feedback on whether the guess was correct or incorrect.

from their partner. The matcher saw their partner gesturing onscreen, unmirrored, and was presented with a matching array of four sentence pairs, as in the training stage. Each sentence pair was labelled with a number from one to four depending on its position on the screen. The position of each pair was randomised at each trial. The matcher could make a selection by pressing the number corresponding to their chosen sentence pair. The experimental design allowed for interruption, so that either the director could terminate recording and streaming by pressing space bar, or the matcher could do so by making a selection. As in training, the target pair was highlighted in green once a selection had been made, with the matcher's selection highlighted in red if it did not match the target. For both participants, feedback was presented for five seconds.

4.2.4 Aims and predictions

The aims of the present experiment were to investigate how participants describe who does what to whom, when the meanings they must communicate are potentially ambiguous as to the participant roles in the event. Using proper names rather than noun phrases, I have discouraged participants from relying on iconic representations of individual arguments and must rely on alternative strategies, such as word order, symbols or indices for participants in the scene, or the possible use of space. I hypothesise that the visual nature of the manual modality will make the use of iconic spatial distinctions likely, though it is possible participants may use other strategies. Furthermore, I expect that particular strategies will be systematised through the use of the signals in interaction and through transmission of the signals to new learners.

4.3 Results

4.3.1 Gesture coding scheme

Gestures produced by participants were coded for the presence of agent, goal and verb. The presence of an agent or goal was coded if any agent or goal could be inferred from a gesture (i.e. not necessarily a particular agent), meaning any gesture that stands in for an agent, or any body-as-subject gesture, where the participant mimes an action as if their body was enacting that motion. For example, in a gesture where the participant pantomimes running as if they were, in fact, running, allows the inference of the participant's body as the agent of the sentence. The verb was coded as a gesture pertaining to the meaning represented by the verb in the target sentence. For instance, the body-as-subject gesture described above would be coded as having the presence of the verb as well as the agent, as both agent and verb can be inferred from the simultaneous gesture.

If an agent, goal or verb was present in a trial video, I also coded the type of gesture used to represent each element. For agents and goals, these types were coded in four ways:

- **Body-as-subject** The participant pantomimes the sentence as if they are the agent (e.g. a pantomimed running action using the gesturer's body as if they were running).

- **Descriptive** Reference to the agent is made through the form of a particular gesture. For example, the use of a single figure to denote a participant in the scene. A different code was given to each single form a participant produced (for example, contrasting a gesture where one index finger is raised with one where the index and middle fingers are raised).

- **Indexical** Participants in the scenes are indexed by pointing or otherwise indicating a location in space as a participant in the scene.

- **Directional** The agent is inferred from the starting direction of a gesture which is not a body-as-subject gesture. This can occur as the path of a verb where the path is simultaneous with the action, or a path gesture that is separate from a gesture describing the action.

Gestures for verbs were coded according to a number of descriptive codes that are similar to the handshape coding described in chapter 3. In addition to these types, there were very few instances of gestures that do not fit into these categories that appeared in gestures sequences; these gestures were not coded for.

- **Body** Gestures that are embodied by the participant, such as a pantomime of a running action.
- **Instrument** The gesture represents an object in some way, such as the palm of a hand representing a book.
- **Handling** The gesture represents how an object is handled, such as a gesture representing the fingers holding a pen.
- **Descriptor** The gesture represents an outline of the physical properties of an object, such as size or shape.

In addition to type gestures, gestures were also coded for location of a gesture and direction of movement. Gestures for agent, goal and verb were coded for a location on the horizontal axis parallel to the participant (either centre, left, or right) and were coded as neutral or non-neutral. Gestures that were coded as verbs were also coded for movement features, such as orientation and direction of movement. Orientation of movement could be along different axes and could go in either direction along an axis (as shown in figure 4.5). Axes here take their orientation labels from Padden et al. (2010), where I term the horizontal axis parallel to the participant the x-axis, the perpendicular horizontal axis is termed the z-axis, and diagonal horiztonal axes are termed the $x + z$-axis, though for the coding scheme, each axis along with its direction was given a unique code (e.g. left to right along the x-axis). I also noted whether or not a directional path separate from the path of the verb gesture was present, and if so, which direction that path followed.

Gestures could be coded as neutral or non-neutral depending on the position of the body (also called "spatially modulated" gestures A. Senghas and Coppola (2001)), or on the location of manual gestures. Figure 4.6 gives examples of neutral and non-neutral gestures for a whole body gesture and manual gestures.

This coding scheme allows the categorisation of gestural data, and can be used to analyse whether participants distinguish between sentence arguments. For instance,

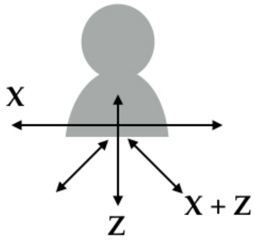
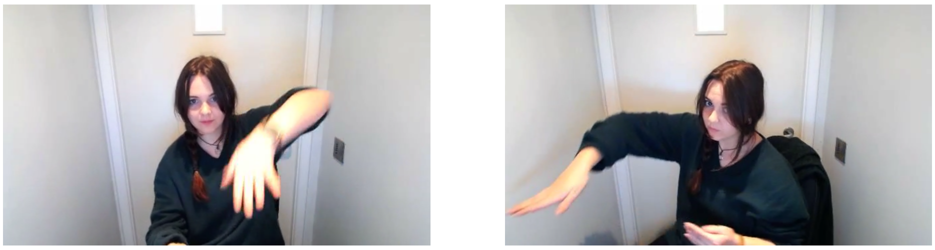


Figure 4.5: Possible movement paths of gestures in the experiment. Gestures were coded for the orientation and direction of movement (e.g. left to right along the horizontal axis parallel to the participant). Labels here refer to the horizontal axis parallel to the participant (x), the horizontal axis perpendicular to the participant (z) and the horizontal diagonal axes ($x + z$), following Padden et al. (2010)



(a)

(b)

Figure 4.6: Gestures showing neutral and non-neutral use of space for body gestures (a) and manual-only gestures (b). In (a), the participant shows a neutral swimming gesture in the left-hand pane, and a non-neutral swimming gesture in the right-hand pane, orienting her body to represent the agent of the target sentence. In (b), the participant in the left-hand pane produces a raised index finger gesture in a neutral, central position, whereas the participant in the right-hand pane produces the same gesture further in front and to his right.

Strategy		
Lexical	**Body**	**Index**
chain 1	chain 2	chain 3
chain 5	chain 4	

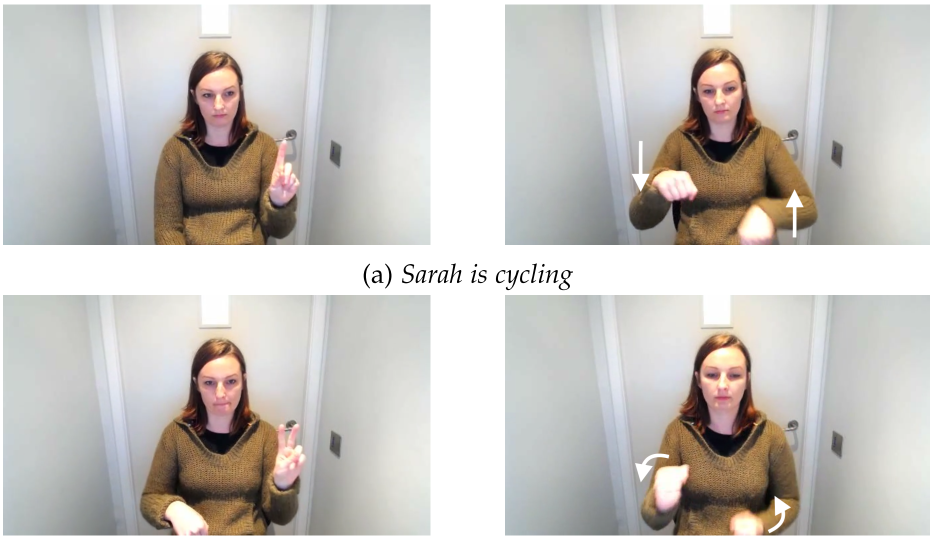
Table 4.2: Transmission chains grouped according to the strategy they use to distinguish between agents.

participants can express arguments differently based on the type of gesture used, or the location and path of the gestures. Therefore, the coding of this data allows the investigation of the the principal aims of this study: 1) do participants distinctly express the agents of the target sentences, and 2) does this distinction become more systematic over generations of learning and use?

4.3.2 Possible pathways for argument distinctions

Both qualitative and quantitative analysis of the data suggests that there are a number of ways in which participants create distinctions between the agents in the target sentences. According to the formal properties of gestures that were coded, participants could distinguish between agents based on the types of agent gestures used, the location of the agent, the location of the verb and the path of the verb. Largely, these formal distinctions can form three groups of strategies: lexical, body and index strategies. Importantly, linguistic use of all of these strategies are attested in natural sign languages (Meir et al., 2007; Padden et al., 2010; Sandler & Lillo-Martin, 2006), though it is the indexing strategy that most closely resembles the spatial grammars that are common across different sign languages (Sandler & Lillo-Martin, 2006; Sutton-Spence & Woll, 1999).

In the following sections, I will describe each strategy in detail, giving examples from gestures produced by participants. Each transmission chain in the experiment utilised one of these strategies in the majority of generations (i.e. 3 out of 5 generations), and can be grouped according to the strategy used (see table 4.2). Through using examples of the strategies used in each chain, I will present an overview of how participants negotiate and develop consistent, systematic distinctions over the course of the experiment.



(a) *Sarah is cycling*

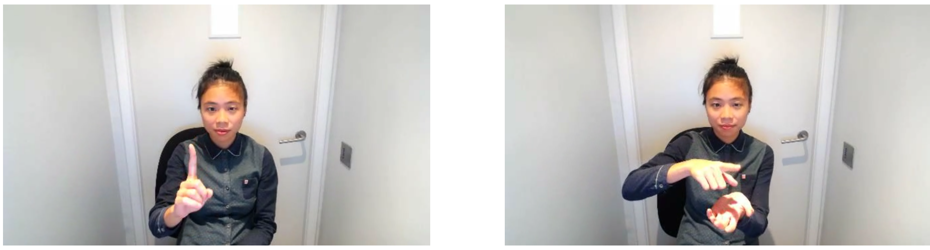
(b) *Hannah is running*

Figure 4.7: Use of 1 and 2 handshapes to represent different agents in a different-agent sentence pair. In this case, 1 is used to denote Sarah, 2 to denote Hannah. Example from chain 5, generation 3.

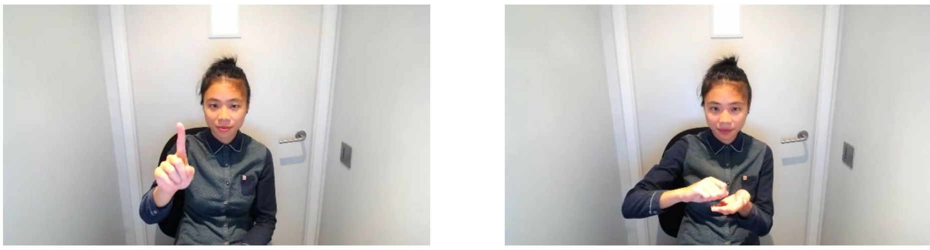
Lexical strategies

Lexical strategies involve the distinction between agents across a pair of sentences being made by the type of gesture used to denote an agent. In both chains in which this strategy is primarily used, chains 1 and 5, participants vary gesture type based on handshape. Figure 4.7 illustrates these handshapes, showing the use of 1- and 2-handshapes, in which either one or two fingers is extended, to denote the difference between agents in the first and second sentences of a pair.

It is possible that these gestures are used to simply note the first and second sentence of a pair, and indeed, there is some evidence that this strategy is used this way initially. However, examples from each chain suggest that these gesture forms eventually distinguish agents rather than sentences. Figure 4.8 illustrates how these forms are used as agent markers rather than discourse markers. In figure 4.8a, showing an example from chain 1, the agent reuses the 1-handshape in gesture sequences for both sentences, noting that the agent is the same in both cases. In figure 4.8b, the 1 and 2 handshapes stand in for agent and goal, respectively, both gestures referring to participants in the target sentence. Further evidence that these gestures are associated with agents rather than sentences arises in a small number of cases where the 1 and 2 handshapes come to stand in for particular agents (for example, one participant uses the 1 handshape to denote Sarah, and the 2 handshape to denote Hannah).

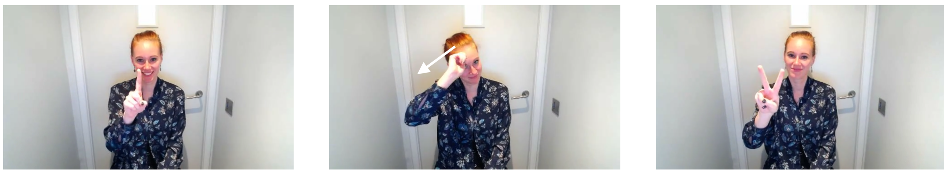


Hannah is kicking a ball to Sarah



Hannah is sending a letter to Sarah

(a)



Hannah is throwing a hat to Sarah

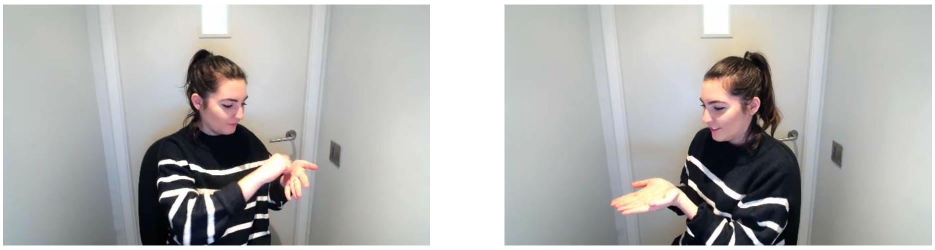
(b)

Figure 4.8: Examples showing the use of 1 and 2 gesture to denote agents. In (a), the participant uses the 1 gesture for both agents in a same-agent trial; here the gesture clearly denotes the agent rather than the sequence of sentences. In (b), the participant uses both 1 and 2 gestures in a gesture sequence, where 1 notes the agent and 2 marks the goal. Gesture examples are taken from generation 5 participants of chain 1 (a) and chain 5 (b).

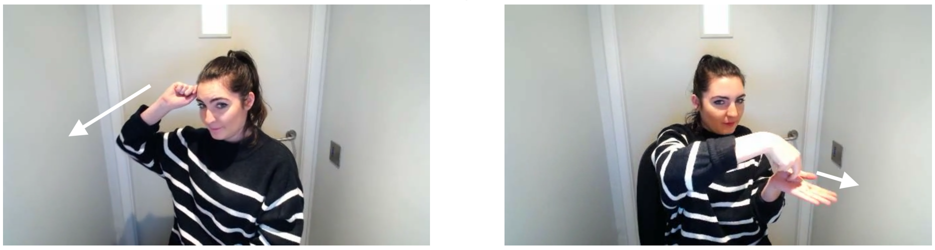
Body strategies

The category of body strategies comprises gestures in which the orientation and position of the participant's body serves to signal differences between arguments in a sentence. Primarily, such a strategy relies on orientation, with participants remaining in the same location but orienting their body to the left or right, providing a contrast between participants on a vertical axis. However, there do exist some instances of this distinction being made purely on body location, where a participant's body is oriented to face the camera straight on, but they move their body to a location either more left-of-centre or more right-of-centre to distinguish between participants in a sentence. The example shown in figure 4.9 illustrates the use of body orientation to denote the agents in a sentence. Here the participant not only varies their body position to denote that the agents in the two sentences of a pair are different, but differentiates arguments consistently, gesturing actions by Sarah in a right-oriented position, and actions by Hannah in a left-oriented position. Use of the body is a highly iconic strategy, relying on actions performed using the participant's whole body, with the agent of a proposition inferred as the participant's whole body, which can take on different roles through the variation of position. Body-as-subject is a phenomenon frequently attested in sign languages, where the signer's body represents the subject of a proposition, or highest ranking thematic role, such as agent or experiencer (Meir et al., 2007). Use of the body in the cases described here is similar, with participants using their own body to represent the agent, leaving the goal of the sentence unspecified.

There is variation among the strategies used by each chain (chains 2 and 4). Though both chains distinguish between agents in different-agent contexts using body position, the extent to which they create clear distinctions in same-agent contexts differs. For example, figure 4.10a illustrates a participant using body position to clearly mark individual agents in same-agent sentence pairs, in which Hannah is gestured with a right-oriented body position even in the same-agent context. Body positions that pertain to particular arguments are used across all contexts in the experiment by this participant, contrasting with the participant from chain 2 shown in 4.10b, where the participant gestures Hannah with a neutral body position, which does not specify who the agent is (i.e. Hannah or Sarah). Indeed, she retains the neutral body position even in same-agent contexts where Sarah is the agent. Importantly, this lack of variation in the same-agent context is not necessarily uninformative; the very lack of variation of body position in the same-agent context contrasts with variation in the different-agent context, allowing their partner to recognise the context and thus narrow down possible correct interpretations.



Sarah is sending a letter to Hannah
Hannah is giving a book to Sarah



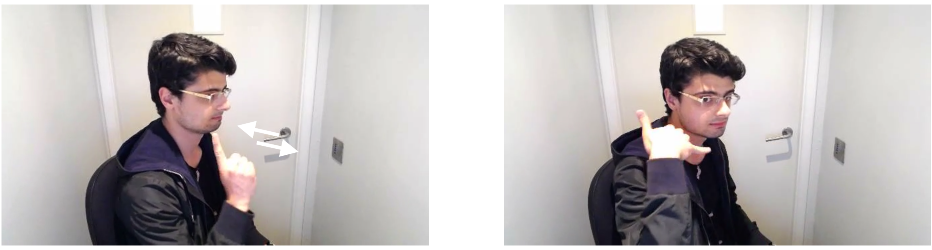
Hannah is throwing a hat to Sarah
Sarah kicking a ball to Hannah

Figure 4.9: Use of body position to denote agents in a target sentence. In both sentence pairs, the participant demonstrates the difference between agents by body orientation. Furthermore, body position denotes particular agents; orientation to the right signals Sarah, orientation to the left signals Hannah. These examples were produced by a generation 5 participant in chain 2.

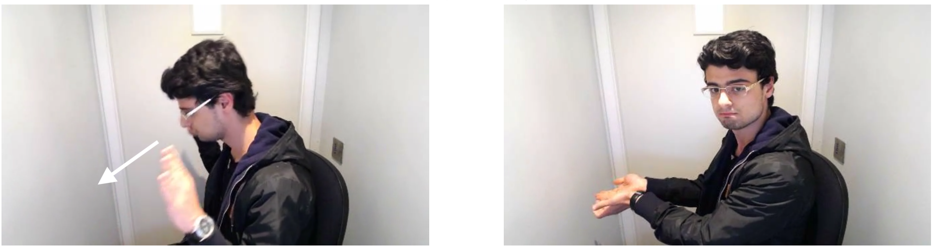
Index strategies

Only one chain in the experiment (chain 3) uses a strategy that relies mainly on indexing. With the indexing strategy, participants differentiate agents in a sentence pair by indexing locations in the gesture space to represent them, primarily with a deictic point (see example shown in figure 4.11). Indexed locations are usually opposite to each other, but this opposition can be set up on the z-axis (figure 4.11a), or on the x-axis (figure 4.11b). Indexing in chain 3 is initially innovated along the z-axis, but develops through generations to the x-axis; it is possible that the parallel axis, from the perspective of a matcher, shows the distinction between agent and goal more clearly than the perpendicular axis.

As with the other strategies described above, there is evidence that indexing strategies become associated with specific arguments in a target sentence rather than simply with whether they are the agent in the first or second sentence of a trial pair. Figure 4.12 illustrates the gestures a participant produces for two sentence pairs, where the index locations for Hannah and Sarah are consistent across the two pairs, suggesting that the indexed locations mark specific agents and goals rather than just the role itself. With the indexing strategy, there is also the possibility of agreement between the verb gesture and the location of the agent, the location of goal, or both

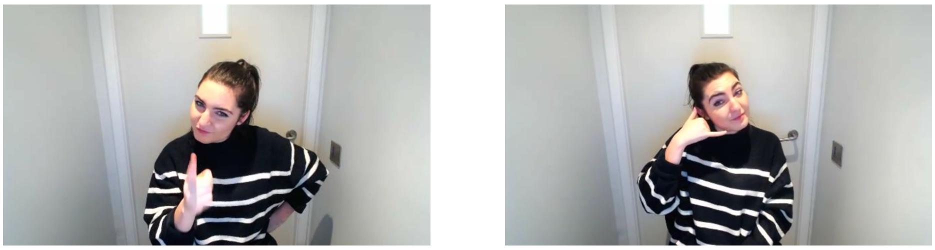


Hannah is scolding Sarah
Hannah is phoning Sarah



Sarah is praising Hannah
Sarah is helping Hannah

(a)



Hannah is scolding Sarah
Hannah is phoning Sarah

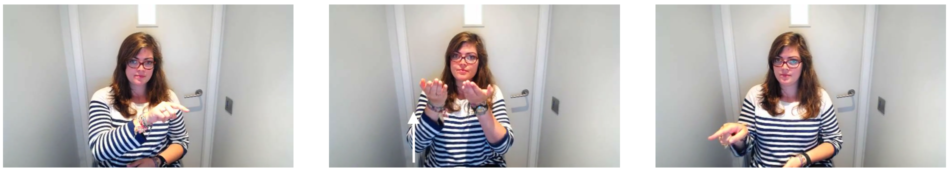
(b)

Figure 4.10: Use of body position to denote agents in same-agent sentence pairs. The participant in (a) uses body position to represent Hannah and Sarah, even in same-agent sentence pairs (orientation to the right denotes Hannah, orientation to the left denotes Sarah). In (b), the same participant that uses body position to distinguish between Hannah and Sarah in the different-agent pairs shown in 4.9 no longer contrasts body position in same-agent pairs, and uses a neutral body position. Examples shown here are taken from generation 5 participants from chain 4 (a) and chain 2 (b).

agent and goal. For instance, figure 4.13b shows a gesture in which the predicate gesture for *giving a book* is directed from the location of the agent to the location in which the goal is indexed. This contrasts with the gesture shown in figure 4.13a, where the direction of movement for the action (here *throwing a hat*) signals no relationship between the locations indexed for either agent or goal. Frequently, gestures that are grounded in the body, such as a pantomime of walking, which include movements of the whole body, show agreement between indexed locations less often than

Sarah is helping Hannah

(a)



Sarah is helping Hannah

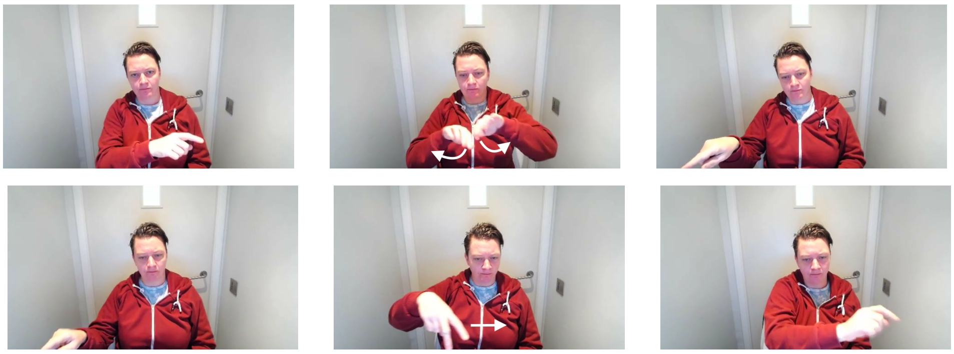
(b)

Figure 4.11: Examples of gestures using the indexing strategy, where separate locations are indexed to represent different arguments from the target sentence. (a) shows a participants using indexed locations on the z-axis. (b) shows a later generation, where the reference axis has moved, and predicate arguments are indexed on the x-axis. The participants shown here participated in chain 3, generation 2 (a) and generation 3 (b).

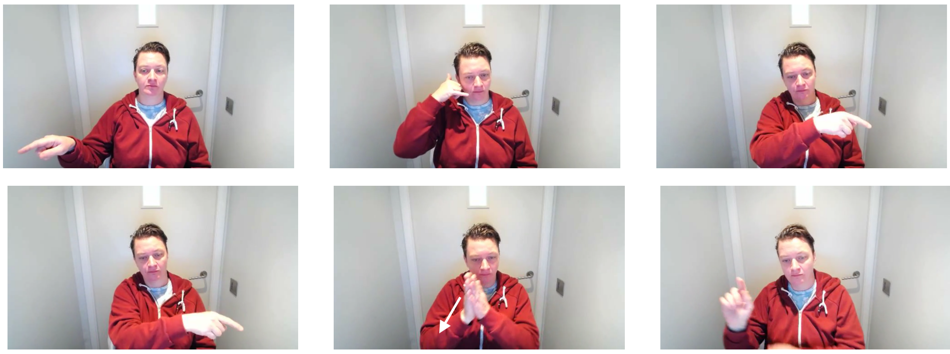
movements that use smaller joints, or do not use the whole body, suggesting that gestures that are less grounded in the body, are more amenable to showing agreement between indexed locations.

Interim summary

Different chains in the experiment rely on different primary strategies to distinguish between agents in the target sentences they have to communicate. However, there seems to be a restricted number of strategies they employ, using only three clear types, all of which use devices attested in natural sign languages, with chains converging on similar strategies. Furthermore, even between different distinguishing strategies, commonalities arise. Participants use their chosen strategy to distinguish agents in different-agent contexts, marking the role of agent, and may further create distinctions in same-agent contexts, by using consistent gesture forms (through gesture type or location) to denote arguments in a target sentence. Such cases present evidence that these forms are becoming reanalysed from simply denoting the role of agent to represent specific agents in target sentences (either Hannah or Sarah). Participants devise iconic strategies, relying heavily on the iconic representations allowed by their own bodies. However, even when highly iconic strategies are employed, participants use them systematically to signal the arguments in a sentence. In the following sections, I will look more closely at the extent to which participants cre-

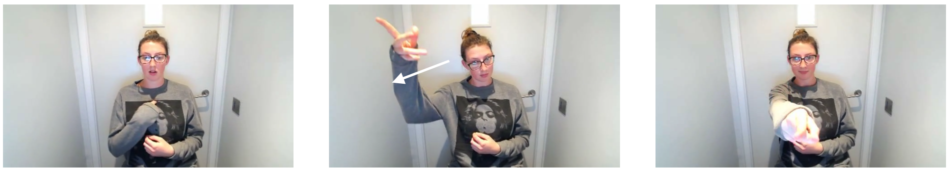


Sarah is swimming to Hannah
Hannah is walking to Sarah



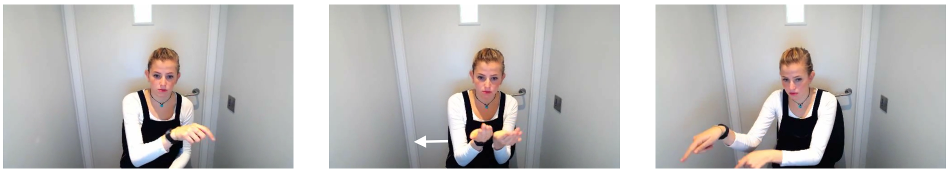
Hannah is phoning Sarah
Sarah is praising Hannah

Figure 4.12: Use of indexing where the index locations are consistent across agents rather than simply signalling that the agents in two sentences are different. In this example, one location represents Hannah (screen left here), and one location represent Sarah (screen right here). This example was taken from chain 3, generation 4.



Hannah is throwing a hat to Sarah

(a)



Hannah is giving a book to Sarah

(b)

Figure 4.13: Examples of indexing gestures where the verb path is neutral with respect to index locations (a) and where the verb path moves between two indexed locations. In (a) the path of the verb does not move between location of the indices denoting Hannah and Sarah. In (b) the direction of the verb agrees with the location of the agent and the goal of the target sentence, Sarah. Examples here are from participants from chain 3, in generation 2 (a) and generation 4 (b).

ate these signals, and how the signals change over generations in the experiment, through the use and transmission of gestures.

4.3.3 The evolution of signalling strategies

The following section will analyse the gestures participants produce to further understand how the different strategies described above lead to systematic formal variations in participants' gestures. As described in section 4.3.2, participants can demonstrate distinctions between agents based on the type of gesture used to describe the agent, the location of the agent gesture, the location of the verb gesture, and the path of the verb gesture. The analyses detailed below will look at differentiation on each property, providing quantitative evidence of the three strategies described above. Gestures for goals in target sentences will not be discussed here: goals were gestured with a specific form in only 29.5% of trials (excluding plain verb trials, where a goal is not encoded in the target sentence).

I analysed whether participants demonstrate a distinction based on the gestural features (agent type, agent location, verb location, verb path) to denote agents in a pair. For the measures described below, I compared the agent in the first sentence of a sentence pair and the agent in the second sentence of a sentence pair, noting whether the agents did or did not show a difference. If agents show no difference within sentence pairs (for example, using body-gestures in neutral position for all trials), this suggests that participants do not create distinctions in their gestures to describe different agents in a sentence. Conversely, if all agents demonstrate a distinction (for instance, iconically representing each event without relation to other events), participants are innovating gestures that represent sentences holistically, and which do not create any regular distinctions between agents in target sentences. If, however, participants demonstrate consistent agent use to different extents, based on whether the target sentences contain the same agents or different agents, the variation of those distinctions would indicate a system in which participants vary gestures denoting different agents in order to distinguish between them.

Proportion of differentiated agents creates a distinction between sentence types

Figure 4.14 illustrates distinctions between agents in a sentence pair in same-agent and different-agent contexts, for the four possible ways of creating such distinctions: by varying agent type, agent location, verb location or verb path. These data corroborate the qualitative analysis that chains converge on particular strategies, as different chains make their distinctions in different ways. For example, chains 1 and 5, categorised as using the lexical strategy in section 4.3.2 here show a difference between sentence types primarily by varying the agent type of their gestures (i.e. using a different form to describe the agent in each sentence). Chains 2 and 5, categorised as using the body strategy, create distinctions on agent location, verb location and verb path. It is important to note that that they show a distinction on three criteria is not necessarily more informative. In the case of the body strategy, agent and verb location and verb path are conflated, as the position of the body signals the position of the agent, but it is also the locus of the verb, and so all three will vary concurrently. In contrast, chain 3 shows its primary difference by varying agent location; using the indexing strategy, participants can use the same gesture type, simply differentiating agents based on their location.

Furthermore, participants' negotiation and eventual convergence on a strategy suggests a trade-off between expressivity and economy in the system. Participants demonstrate contrasts between agents in different sentence types, providing the necessary expressivity to distinguish between different targets. However, they also demonstrate economy, settling on distinctions based on one primary strategy at the expense of other strategies, evidencing an increase in simplicity. For instance, generation 1 of chain 1 shows variation on agent location for two-thirds of different-agent sentence pairs, but reliance on this distinction reduces following generation 2, when a distinction based on agent type takes over. Similarly, chain 3 often differentiates

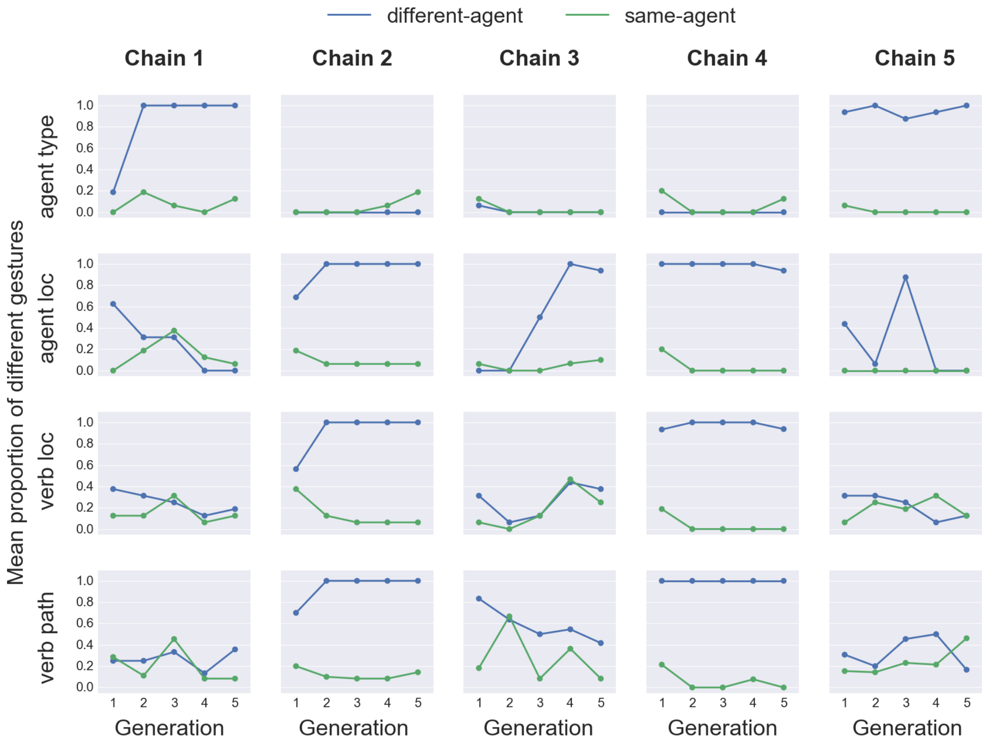


Figure 4.14: Proportion of gestures that differentiate agents in the target sentence, shown for same-agent and different-agent contexts. Rows show the differences based on which element of the gesture sequence varies (agent type, agent location, verb location, verb path), and columns show the values for each chain, at each generation. Blue lines show data for different-agent sentences, green lines show data for same-agent sentences. All chains show differences based on the context in which the agents occur, demonstrating a higher proportion of differences in the different-agent context. However, chains make these distinctions in different ways, that correspond to the strategies discussed in section 4.3.2.

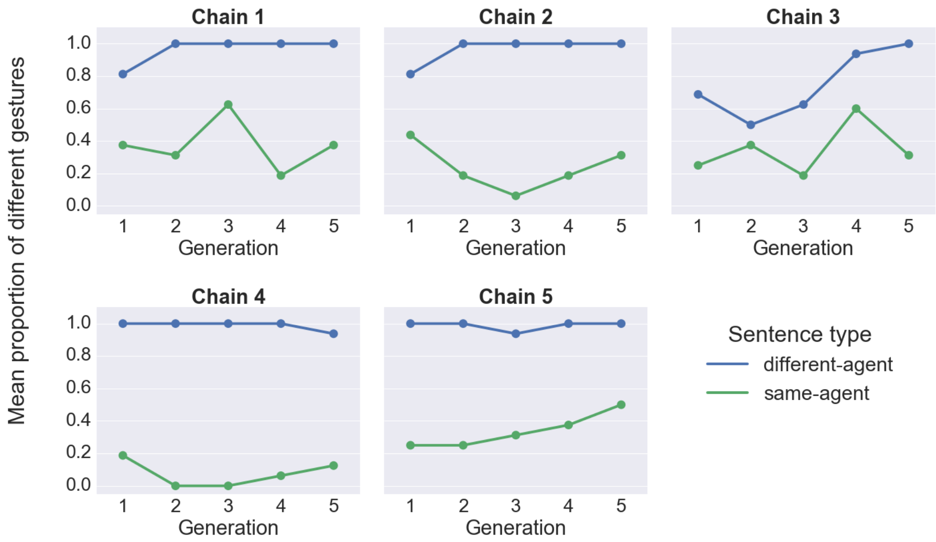


Figure 4.15: Proportion of gestures that differentiate agents in the target sentence, shown for same-agent and different-agent contexts. Blue lines show data for different-agent sentences, green lines show data for same-agent sentences. Here, differences are counted as a binary feature, noted as whether or not the agents are different, based on any of the possible features (agent type, agent location, verb location, verb path) shown in figure 4.14.

between same- and different-agent sentences using verb path in generation 1. As the indexing strategy becomes adopted, through variation based on agent location, the use of distinctions based on verb path reduces over generations.

Though figure 4.14 provides a clear overview of how participants make these distinctions, I can collapse this view over categories to focus the question further, in order to ask whether or not participants make a distinction between agents in target sentences, and whether this distinction is used differently in different-agent and same-agent contexts. For this, the difference measure described above is converted to a binary measure, that notes whether or not a difference is signalled between agents in both sentences of a target pair, regardless of how this distinction is made (i.e. if they differ on any of the features shown in figure 4.14). Figure 4.15 illustrates this measure, indicating that though chains of participants converge on different strategies, all chains differentiate between agents in contexts where potential ambiguities arise.

All chains demonstrate a difference between the extent to which they represent agents differently in different-agent contexts and in same agent-contexts, participants creating distinctions between different agents in target sentence pairs. I analysed the proportion of agent differentiation using a logistic mixed effects regression, implemented with the lme4 library (Bates et al., 2015) and R (R Core Development

Team, 2008). The model included sentence type (with same-agent type as the intercept) and generation as fixed effects, as well as their interaction. I included chain, target and participant as random effects with random intercepts. Random slopes of generation and verb type were included for chain and target, respectively, and the random effects structure for participant was nested within chains. Model comparison revealed a significantly better fit for the model described here compared to a reduced model without the interaction term ($\chi^2 = 11.51, p < 0.001$). The model demonstrated a significant effect of sentence type (different-agent context: $\beta = 2.73, SE = 0.46, z = 5.91, p < 0.001$), but no significant effect of generation for the same-agent sentence type ($\beta = 0.04, SE = 0.12, z = 0.29, p = 0.77$). However, the model demonstrated a significant interaction between sentence type and generation ($\beta = 0.69, SE = 0.22, z = 3.12, p = 0.002$). Participants demonstrate significantly different strategies for signalling agents in different-agent contexts versus same-agent contexts. Furthermore, this contrast develops over generations, with a general trend for further divergence between contexts.

Use of non-neutral space emerges early and depends on differentiation strategy

I noted the use of neutral and non-neutral space in participants' gestures. Though the use of space can correlate with the distinction strategies participants use (e.g. use of indexing gestures may entail the use of non-neutral space), the measure accounts for how frequently participants perform gestures in non-neutral space, as placing referent signs outside the usual neutral signing space is often indicative in natural sign languages of the instrumentation of a spatial grammar (A. Senghas & Coppola, 2001). The proportion of gestures performed in non-neutral space is shown in figure 4.16, for generations in each chain, and shown for each sentence type. Non-neutral space is used to some extent by all chains, though there are opposing trends. In chains 1 and 5, which use the lexical strategy, they exhibit a reduction in the use of non-neutral space, which becomes redundant as specific gesture forms are conventionalised. In the remaining chains, participants use non-neutral space to differentiate agents, both across all contexts and in different-agent contexts only. Interestingly, the use of non-neutral space is not used contrastively. Rather, the contrast relies specifically on the parameters that participants vary (i.e. the particular differentiation strategy for a chain). For instance, chain 3 demonstrates a contrast between different-agent and same-agent contexts in the use of the indices that denote different agents, but do not contrast the use of neutral and non-neutral space; all indices are placed in non-neutral space. It is the contrast of different non-neutral locations that differentiate agents, not the the use of non-neutral space itself. This points to the iconicity of

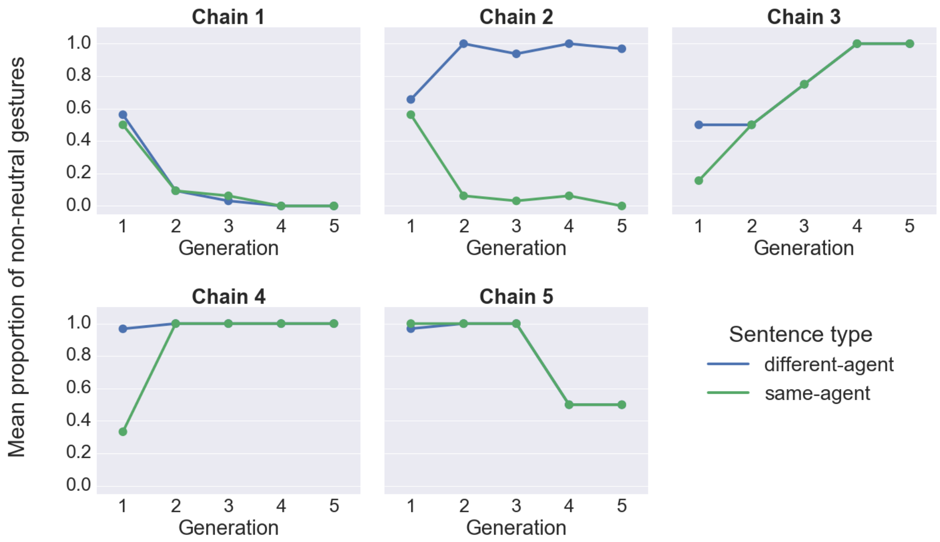


Figure 4.16: Proportion of gestures performed in non-neutral space, in different-agent and same-agent contexts. Blue lines represent different-agent contexts, green lines same-agent contexts. Individual plots are shown for each chain, over all generations of the chain.

strategies participants use. Use of non-neutral space emerges early to denote agents that differ from each other, as the participant's body in neutral space cannot contrast multiple agents. Participants exploit the iconic possibilities of non-neutral space to make the distinction between agents clear, maximally contrasting agent locations in non-neutral space, rather than contrasting a non-neutral location with a relatively ambiguous neutral location.

Verb type does not effect agent distinctions

Finally, I analysed the proportion of gestures that differentiate agents in a sentence pair, based on the four verb types that participants were presented with in target sentences (plain verb of motion, verb of motion with goal, physical transfer, non-physical transfer). Spatial grammar in sign languages is not implemented wholesale in predicate constructions; it serves to disambiguate agents and goal participants in complex transfer verbs, such as *ask*, or *telephone* (though there is cross-linguistic variation concerning exactly which verbs take spatial agreement).

Figure 4.17 illustrates the proportion of different agents gestures for each verb type. A logistic mixed effects analysis investigated the effect of verb type on the proportion of differentiated agent gestures in a sentence pair. Verb type was included as the fixed effect, with chain, target and participant included as random effects with random intercepts. Random slopes of generation and verb type were included for chain and target respectively, and the random effects structure for participants

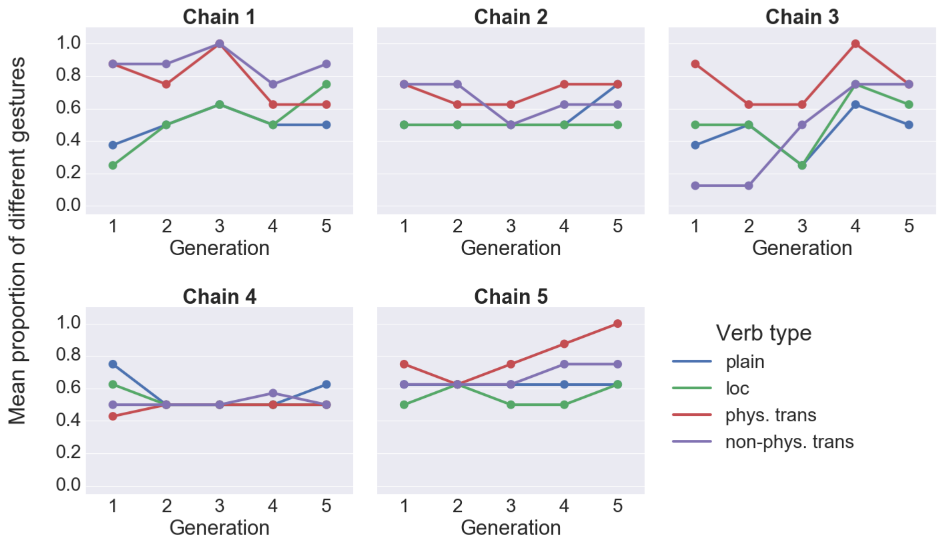


Figure 4.17: Proportion of gestures that differentiate agents in the target sentence pair, based on verb type (either plain, locative, physical transfer, non-physical transfer, shown with coloured lines). Individual plots show the data for each chain, over all generations of the chain.

was nested within chains. The model showed no improvement over the null model ($\chi^2 = 1.82, p = 0.61$), suggesting that participants show no variation in how they differentiate agents in target sentence pairs based on the verbs used in those sentences.

4.3.4 Results summary

Participants in the experiment produce gestures that differentiate between agents in the target sentences, and chains of participants that interact with and learn from each other converge on particular strategies to create these distinctions. In the present data, I identified three broad strategies: the lexical strategy, the body strategy, and the index strategy. All three strategies rely on iconic representations of the target sentences, but these representations become regularised and systematised to consistently distinguish between agents in target pairs. Participants demonstrate contrasts between discourse contexts in which the agent of two sentences is different (different-agent) and in which the agent of two sentences is the same (same-agent), producing more differentiated forms in different-agent contexts than in same-agent contexts. Participants demonstrate use of non-neutral space in their gestures, but the contrast between neutral and non-neutral space does not provide a distinction between agents in most chains; it is the mechanisms that define the strategies described above that create contrasts between agents and between sentence types. Finally, these findings suggest that participants did not create distinctions to different extents based on the verb in the target sentences. Strategies used to differentiate agents were used across the whole system of gestures.

4.4 Discussion

Participants in the experiment use iconic strategies to refer to agents in the target sentences, and do so from early generations in transmission chains. However, the strategies that participants use to distinguish between agents demonstrate systematisation and convergence as participants interact and learn from each other. Here, I will discuss how the strategies participants use to convey distinctions map onto features of natural sign languages, and how the development of these systems over time mirrors data from emerging sign languages.

4.4.1 Differentiation mechanisms in natural sign languages

The lexical strategies used by chains 1 and 5 do not make contrastive use of space in the experiment, but do contrast agents in target sentences using distinct gestural forms for different agents. It is unsurprising that some participants relied on what could be considered a more lexical contrast. All languages are able to use lexical forms to distinguish people or objects, and sign languages are no exception.

The gestural forms used in both cases are grammaticalised number handshapes (holding up 1 or 2 fingers), which developed from gestures that simply indicate the sentence in a target pair (first or second). These forms eventually stand in for individual names, and possibly act as pronominal forms for the proper names of Hannah and Sarah. Importantly, it appears that use of these lexical forms contrasts with the use of space; spatial contrasts are redundant in this case. Certainly, spatial modulation is not the only mechanism available in sign languages to mark participant roles (Lillo-Martin & Meier, 2011; Meir et al., 2007).

The remaining three chains show some use of spatial modulation to distinguish between agents in target sentences. The body strategies exemplified by chains 2 and 4 involve movement of the participant's body to positions that distinguish between agents in target sentence; for the indexing strategy used by chain 3, participants' bodies remain in neutral position, but indices referencing agents and goals in the target sentences contrast in position. The indexing strategy used by chain 3 could be considered to be the closest parallel to the grammatical use of space in sign languages, where deictic points index referents in space and the paths of agreement verbs move between indexed referents. However, both indexing and body shift devices are attested in sign languages, where indexing is common in agreement verb constructions, and body-shift occurs in role shift contexts (Cormier et al., 2015; Kocab et al., 2014). Though role-shift is not the same linguistic phenomenon as person marking in verbal constructions, the relationship between pronominal forms and body position in role-shift contexts indicates that these mechanisms are not unrelated to each other (Cormier et al., 2015), and indeed, body-shift can be used to differentiate predicate arguments in ASL (Padden, 1990). Furthermore, use of the body to represent the agent, or the highest thematic role in a proposition, is proposed by Meir et al. (2007) to hold a privileged position in sign language, relying on the iconic property of the body's agency to represent animate agents. Across all chains in the experiments, participants show iconic use of the body to represent predicates in target sentences, for instance with a running gesture, using body-anchored gestures to iconically represent the action. I discuss the iconicity of body-anchored gestures further below, in section 4.4.2.

The index strategy that most closely resembles typical spatial reference in sign languages also demonstrates development over generations that follows a trajectory similar to that found in emerging sign languages. The implementation of this strategy in chain 3 begins on the z-axis, with participants in generation 1 pointing at themselves for the agent, and indexing space directly in front of them for the goal of the target sentence (see figure 4.11b for an example of this). Here, the participant's body is still being iconically used to represent agency. By generation 3, however, indexing has been abstracted away from the body to the horizontal x-axis. This change reflects changes noted in two young sign languages, ABSL and ISL (Padden et al., 2010). Signers in both ABSL and ISL show a preference for verb paths that are oriented with respect their own bodies, on the z-axis, following the body-as-subject preference posited by (Meir et al., 2007). However, signers from more recent generations of both languages show a decrease in the use of z-axis paths compared to older signers, and an increase in the use of abstracted verb paths on the x-axis, reflecting the change in chain 3 from z-axis indexing to x-axis indexing. Padden et al. (2010) and Meir et al. (2007) suggest that the two axes represent competing iconicities: one which represents animate agents with the most accessible resource available, the signer's body, and one which is able to iconically represent non-first person agents in contrast to each other, but which requires abstraction away from the body. High levels of iconicity are prevalent in the gestures participants produce; body-situated gestures are used throughout the experiment to different extents. As such, the problem of competing iconicities, and iconicity in general, warrants its own discussion.

4.4.2 Iconic representation through use of the body

All participants make use of iconic body representation to gesture verbs throughout the experiment (though they differ otherwise in whether they use the body to make contrasts between target sentence meanings). The body has been suggested to be a fundamental device for representation in sign languages (Taub, 2001), but particularly has been highlighted as an elementary form in the representation of predicate forms (Meir et al., 2007), to the extent that sign languages have been suggested to contrast only two grammatical person categories: first person (articulated on the body) and non-first person (articulated in space around the signer) (Lillo-Martin & Meier, 2011). Forms that frame an event in relation to the body (i.e. using the z-axis) appear before forms that abstract away from the body (Meir et al., 2007; Padden et al., 2010), and even in older sign languages such as BSL, signers evidence a preference for body-anchored verb paths (Cormier et al., 2015). Even in languages, such as ASL, that demonstrate higher levels of third person to third person reference along the x-axis, there are still sets of verbs that anchor the verb form to the body, such that the path from agent to goal must originate from the signers body, even for a third-person referent (Liddell, 2003; Meir et al., 2007). As such, the prevalence of body-use in the present experiment supports the primacy of the body in the gestural modality and points to the iconic advantage of a visual modality for representing animate agents.

The present findings do not demonstrate any differences between verb types, with similar strategies used to represent agents across all four categories. Different verbs types were included in the study to test whether different semantic relations lead to different gestural representations; for example, that spatial verbs may be the locus of change for gestures in the system because they are more iconic and therefore would lend themselves to iconic spatial mappings. It is possible that this lack of difference is a result of the experimental design, as participants are presented with all verb types equally across blocks, which may encourage them to use the same strategy in all contexts. However, that we find no difference between verb types is consistent with data from ABSL and ISL (Padden et al., 2010), indicating that signers in two young sign languages use similar strategies independent of verb type, particularly in early generations of both languages. From this data, we cannot pinpoint whether abstraction of iconic strategies to arbitrary use of space is dependent on particular verb constructions.

Participants throughout the experiment demonstrate a reliance on iconic representation, particularly with respect to use of the body, where participants embody agency in verb gestures. However, the movement in chain 3 from the more iconic body-situated gestures to abstracted indices placed on the x-axis is indicative of an interaction between iconicity and systematic structure over time. Previous research in both natural language and experimental contexts has indicated that the development of linguistic structure over time leads to a reduction in iconicity (Frishberg, 1975; Theisen et al., 2010; Theisen-White et al., 2011), and the interaction between iconicity and systematicity has been suggested to play a particularly important role in the specific context of verb agreement. Older sign languages such as ASL have a greater prevalence of arbitrary spatial reference than younger sign languages, such as ABSL and ISL, which demonstrate highly iconic body-as-subject strategies (Padden et al., 2010). A remaining concern then, perhaps, is why more chains in the present study did not converge on the indexing strategy, which appears to reflect the most common verb-agreement device for sign languages cross-linguistically. Principally, this experiment models the development of person marking in emerging gestural systems, and evidence from natural emerging sign languages suggests that the grammatical use of space does not emerge instantaneously, but takes time to develop (Kocab et al., 2014; Lillo-Martin & Meier, 2011; Meir et al., 2007; Padden et al., 2010). The findings in the present chapter support the natural language data, suggesting the convergence of a linguistic community on a differentiation strategy and subsequent systematisation of that strategy over generations in a chain.

However, of relevance here is that the prevalence of arbitrary spatial grammars encoded linguistically across sign languages does not remain uncontroversial. Data elicited from BSL signers suggest that body-situated verb agreement on the z-axis is more common than previously thought (Cormier et al., 2015), particularly in comparison with agreement on the x-axis. Furthermore, recent theoretical considerations of ASL, a language that demonstrates consistent use of x-axis agreement between third person referents, suggests that agreement or person-marking systems are not fully linguistic but combine linguistic verb morphemes with deictic gestural points specifying referent location (Liddell, 2003; Lillo-Martin & Meier, 2011)[1]. Given the iconic nature of verb agreement systems cross-linguistically in sign languages, the iconicity evidenced by participants in the present study serves to reflect the importance of iconicity in bootstrapping a communication system, and the development of these systems over generations further corroborates findings that iconicity reduces as signals become increasingly systematised.

Participants throughout the experiment demonstrate the use of iconic gestures moderated by the increasing regularity of the system, such that the iconicity inherent in participants' gestures is not unconstrained. For example, participants use of

[1]Lillo-Martin and Meier (2011) liken ASL agreement verbs to pronouns in spoken language that are combined with co-speech gesture; a referential index acting as a pronoun combines simultaneously with a deictic point specifying a referential location for the index

the body in chains 2 and 4 is moderated by their use of body position to distinguish between agents. Though the distinction is enacted through embodiment of the agent, and is thus iconic, there is nothing iconic about positioning particular agents in particular positions. No locative information is given in the target sentences that participants must convey that could iconically position Hannah or Sarah in topological space; participants' decisions to position either agent to the left or to the right are essentially arbitrary. However, for all chains that contrast space (chains 2, 3 and 4), there are consistencies in how they contrast agent gestures, in that participants tend to position arguments in maximally contrastive positions, such as left versus right, or close to the gesturer's body contrasted with the furthest visible position away from the body on the z-axis. Maximal contrasts, though still somewhat iconic, do not provide a direct iconic mapping from target meaning to gesture, as the maximal space between gestures does not convey differences in the topological space between Hannah and Sarah in the target sentences. For example, one might assume that Hannah and Sarah are separated by a greater distance in the sentence *Hannah is phoning Sarah* than in the sentence *Hannah is kicking a ball to Sarah*, yet these spatial differences are not iconically represented in the gestures produced by participants. The use of maximal contrasts is a reflection of agreement verb systems in natural sign languages (Lillo-Martin & Meier, 2011), and is indicative of spatial mappings becoming systematised to represent the thematic roles of agent and goal.

4.4.3 The evolutionary pathway for spatial grammar

One aim of the present study was to investigate the role of cultural evolutionary processes in the development of spatial grammars. Evidence from natural sign languages suggests that the systematic use of space does not emerge fully-formed in a language, but takes time to develop (Lillo-Martin & Meier, 2011; Meir et al., 2007; Padden et al., 2010). The findings here support a gradual evolution of the use of space. Participants demonstrate a difference in the first generation of their transmission chains between different sentence types. However, the mechanisms they use to create those distinctions are not consistently employed early on, but take time for participants to converge on consistent strategies that further systematise. Participants do not innovate new strategies at each generation but converge onto the same strategy used by their model. Further change after convergence suggests the entrenchment of these strategies and further regularisation of forms using these strategies (for instance, if a particular position becomes associated with Hannah rather than Sarah). Interestingly, the most gradual changes appear in chain 3, which saw the abstraction of their indexing system away from the body, so that verb paths travelled on the x-axis instead of the *z*-axis. Contrasted with the highly iconic body strategies used by chains 2 and 4, the ability to abstract indexed forms away from the body points to the interaction between iconicity and systematicity, such that a move away from reliance on iconic strategies allows the further systematic development of participants' gestures, that is not evinced clearly in chains 2 and 4.

The proportion of distinctive gestures increases across generations, and systematisation of these gestures is further evidenced through increased divergence between sentence contexts, suggesting that particular forms become increasingly associated with particular contexts and thus demonstrating increased consistency found in previous studies using transmission chain experiments (Carr et al., 2016; Smith & Wonnacott, 2010). Furthermore, there is evidence of the trade-off between expressivity and compressibility that earlier work has asserted plays a crucial role in the evolution of linguistic structure (Carr et al., 2016; Culbertson & Kirby, 2016; Kirby et al., 2015, as well as chapter 2 of this thesis). Participants demonstrate signals that are expressive enough to be comprehended by their partner, whilst minimising the number of strategies they must use to convey the target sentence meanings. In several chains at generation 1, participants differentiate agents in target sentences in different ways (for example, in chain 1 they use agent type, agent location and verb location in generation 1), but as they converge on one strategy, the use of distinctions based on other mechanisms reduces. No chain that uses gesture-type distinctions makes use of contrastive space; indeed, the use of non-neutral space reduces over generations for chains 1 and 5 (see 4.16). Similarly, no chain that employs a spatial strategy (either body or index) makes a distinction on the form of the gestures for agents or goals. Finally, this trade-off is most clearly demonstrated in the difference participants make between sentence types. Participants consistently contrast event participants in different-agent contexts, but not in same-agent contexts; they do not represent each target sentence event holistically but express similarities across target sentences with similar gestural forms.

4.4.4 The limits of space

Participants in the present experiment create systems of gestures that distinguish agents and goals in target events. They rely on iconic strategies to encode these distinctions, using mechanisms (such as body-shift and indexing referent locations) that are found in natural sign languages. As participants communicate and learn from each other, they converge on a consistent mechanism for creating these increasingly distinct gestures to describe agents in target sentences. However, gestures do remain highly iconic and the rate of evolution of these systems following convergence on a strategy is low. I suggest there are several possible explanations. Firstly, the use of any spatial mechanism will show a high reliance on iconicity; it may prove difficult to move towards arbitrary gestures for animate agents when the participants' bodies afford a simple, iconic representation for such agents. The slow rate of evolution of a highly iconic gestural system would corroborate previous experimental findings that suggest that the availability of iconic mappings inhibits the emergence of systematic structure (G. Roberts et al., 2015; Verhoef et al., 2015). Secondly, the gestural properties inherent even in fully developed verb agreement systems, such that they make use of deictic points and iconic spatial mappings that may be inconsistently realised across discourse contexts, cannot be ignored (Cormier et al., 2015; Liddell, 2003; Lillo-Martin & Meier, 2011). There are limitations within natural languages on how arbitrary such spatial mappings can become, and these same restrictions may apply here.

For example, Padden (1986) described instances of body-shift in ASL that serve to denote different arguments of a sentence, such that one body position denotes the agent and another contrasting position denotes the patient, termed *contrastive role-shift*. However, Padden (1986) argues that there are physical limitations on the number of roles this type of mechanism can contrast - two at the most. Similarly, there are possible limits to how much a system that relies on indexed referents can be systematised. As Liddell (2003) points out, reference points can occur at any location in the continuous space around the signer, and therefore provide an uncountably large set of possible locations to index, and re-index, across discourse situations. It proves somewhat difficult to systematically categorise a continuous space when there are few physical constraints on how that space can be used. The iconic possibilities of spatial reference perhaps limit such a mechanism's systematisation. The target sentences presented in this experiment only contrast two participants, and as such do not allow us to fully test the limitations of spatial reference. Future investigation, therefore, would open the possibility to study even more complex sentence frames, perhaps requiring multiple (i.e. more than two) sentence participants that interact across discourse contexts.

The study detailed in this chapter has expanded the experimental cultural evolutionary framework applied thus far in this thesis to investigate the effects of interaction and transmission on a complex morphosyntactic phenomenon, the likes of which has not yet been examined under a cultural evolutionary framework. Furthermore, the study focuses on the systematisation of spatial reference, a feature considered to be heavily modality-dependent, and which has lacked experimental investigation as a result. The findings from my investigation support data from emerging sign languages such as ABSL and ISL, which suggest that systematic spatial reference does not emerge fully-formed in a language. Such complex constructions take time to develop and become systematised away from more iconic representation, to representations that exhibit lower iconicity and increased consistency across discourse contexts. Corroborating the findings of Padden et al. (2010), I find no effect of verb type on the emergence of spatial modulation; strategies used to distinguish agents in target sentences are applied across the full set of verbs. In summary, this study points to the reliance by participants on iconic strategies that are systematised through interaction and transmission, but also suggests that there are limits to that systematisation that directly relate to the modality-specific effects of iconicity.

Chapter 5

Summary and conclusions

5.1 Aims and contributions of the thesis

In this thesis I aimed to elucidate the roles that the domain-general cultural evolutionary processes of interaction and transmission play in the evolution of systematic linguistic structure. I built on previous work in the cultural evolution of language to more precisely understand how interaction between language users and transmission of a language to new users serves to shape linguistic structure. The experiments I have described here investigate how communities change communicative signals, through interaction and transmission, and contrast interaction and transmission to understand their particular contributions.

I have presented a novel methodology, artificial sign language learning, to complement research into emerging sign languages. As such, a secondary goal of this thesis was to bridge the gap between experimental research on the evolution of language and natural language research that is valuable to our understanding of newly created languages. By modelling the evolution of linguistic structure in the manual modality, I have examined a modality with linguistic potential, to analyse the creation of communicative signals by experiment participants, and understand the modality-general evolution of linguistic structure. The use of the manual modality has also allowed me to examine modality-specific changes in the communication systems my participants produce, to understand how the demands of a physical modality influence the structure of language.

The experimental work presented in chapter 2 contrasted interaction and transmission separately, and in combination with each other, to demonstrate the precise roles that each mechanism plays in the emergence of systematic structure. The findings in this chapter indicated that both interaction and transmission are required to drive the evolution of systematic and communicatively efficient systems; neither interaction alone nor transmission alone is sufficient for systematicity to evolve. Interaction provides a pressure for communicative efficiency, leading, in isolation, to symbolic systems that minimise production effort whilst maximising communicative accuracy. In contrast, transmission alone leads to systematic, learnable systems, but ones which demonstrate redundancies and which lack efficiency. The combination of both interaction and transmission led to systems in which parts of gesture sequences were re-used to signal similarities in meanings, but in which redundancies in gesture sequences were reduced. The study described in this chapter corroborates previous work that examined interaction and transmission in combination (Carr et al., 2016; Kirby et al., 2015), but further clarifies the roles that each mechanism plays in the emergence of systematic structure.

In chapter 3, I detailed an experiment that analysed how communication between pairs of participants affects the emergence of systematic category distinctions, compared to the signals produced by individual participants. Findings from this study suggested that natural, iconic representations are used by individual gesturers to signal objects and events, though iconicity does not systematically distinguish between categories of events at this stage. Communication leads to further development of these systems, through alignment between participants and convergence on a shared system. Gestures become more expressive, with participants producing more individual elements in sequences of gestures to describe target scenes. The gestures participants produced evidenced the seeds of category distinctions: participants varied signal order and basehand use across scene types. However, participants heavily relied on iconic representations that showed more variation between target items than between scene types, underscoring the inability of interaction alone to lead to systematic linguistic structure. Whilst still playing a role in changing the forms participants use, interaction is insufficient for producing systematic distinctions along scene types, further supporting the results found in chapter 2. The experiment described in this chapter also demonstrated how artificial sign language learning can be used in comparison with data from emerging sign languages, using a method which had previously been used to study the productions of homesigners and signers of an emerging sign language, Nicaraguan Sign Language (NSL). The results I presented are comparable and consistent with the results described by Abner et al. (2016) and Abner et al. (2015), providing a direct link between experimental and field research.

Chapter 4 examined the cultural evolution of a complex linguistic construction, spatial agreement. This work expanded on previous experimental work that has implemented artificial language learning paradigms (Carr et al., 2016; Kirby et al., 2008, 2015; Silvey et al., 2014), using a meaning space of complex events, rather than individual concepts or items, to test how interaction and transmission in combination can lead to the emergence of complex morpho-syntactic structures. Furthermore, the iconic affordances of the manual-visual modality allowed iconic representations of sentence arguments in space, and thus provided fertile ground for studying the effects of iconic representation on the emergence of systematicity. Participants produced gestures that systematically distinguished arguments in target sentences, and frequently made use of iconic space to make those distinctions. The iconic strategies participants relied on were systematised through interaction in pairs and transmission to new learners, though the systematisation of these gestures was limited by participants' reliance on iconic representations. These results offer a novel insight into the interaction between modality-general processes that effect systematicity and the modality-specific reliance on iconicity that interferes with the evolution of systematic structure. The insight into this interaction between systematicity and iconicity also contributes to a wider debate in sign language linguistics, about how iconic, gestural elements (such as deictic points) are incorporated into a structured linguistic system. Furthermore, these results reflect findings from emerging sign languages that suggest that spatial agreement does not emerge immediately in a linguistic system, but that it takes time, through use of a language and through transmission of a language over generations, for a consistent spatial agreement system to evolve (Meir et al., 2007; Padden et al., 2010).

The evidence taken from these experiments together serves to elucidate the relationship between individual learners and the populations in which those individuals interact and learn. Though learners demonstrate cognitive biases or natural preferences for particular representations (such as iconic representations), these preferences in the individual do not give rise to the systematic and symbolic structures that characterise human languages. Instead, interaction between learners in a population and the transmission of language to new learners, provide the functional pressures that drive the emergence of linguistic structure. These modality-general processes act across linguistic domains of varying complexity, driving the cultural evolution of structured language from non-linguistic communication. However, these processes may interact with modality-specific forces that further adapt linguistic systems to a physical modality. As such, the experiments I have presented account for modality-general processes that work across both sign and speech, but also provide a view of the impact of modality on the emergence of language.

The work presented in this thesis has also demonstrated the development of a novel method for studying the cultural evolution of language: artificial sign language learning. The application of this method expands previous experimental work in evolutionary linguistics to the manual modality, reducing interference from participants' existing linguistic knowledge that may be problematic in orthographic or verbal applications of artificial language learning. Artificial sign language learning allows the setup of a controlled, manipulable environment with which to examine the early evolution of language, complementing work on emerging sign languages that offer the only direct available evidence of language creation. Furthermore, the particular experimental setup I detail presents a significant departure from previous silent gesture work. Allowing participants to take part in a controlled space (the experiment booth), with interaction facilitated by video streaming, the experimental setup I have used gives the experimenter greater control and produces clean and focussed video data that only captures participants' productions in the task, facilitating data analysis. The restriction of recordings to trial gestures, eliminating extraneous video footage, allows the videos to be used more easily as training stimuli for new participants, providing a viable platform for experimental models of transmission and learning. Data collection in previous silent gesture studies has necessitated the presence of an experimenter to monitor recording equipment and present stimuli; artificial sign language learning reduces the interference that the physical presence of the experimenter may have.

The artificial sign language methods described here also provide the flexibility for implementation across different linguistic and social contexts, as evidenced by the range of experiments I have described in this thesis. It is possible to manipulate and more precisely control participants' interactions, giving the experimenter control over when participants are able to interact, and for how long, as well as over aspects of interaction, such as the possibility of interruption and feedback. This ease of control over interaction, as well as over learning contexts, allows the experimenter to design conditions that can robustly contrast different cultural evolutionary mechanisms, in addition to individual participant productions. This also offers a wide application of the paradigm, laying the ground for the future work that I discuss in section 5.3.

5.2 Limitations of the present work

This thesis has described the first implementation of a novel methodology, artificial sign language learning, and has demonstrated its efficacy in probing questions about the cultural evolution of linguistic structure in the manual modality. However, this thesis has described the first attempts to apply this method, and has by no means fully clarified the roles of interaction and transmission in language evolution. Indeed, the evidence provided by the present work suggests that the relationship between individuals, the people they interact with and the people they learn from, is a precarious one. The structural differences between communicators' productions in chapter 2 in comparison with those in chapter 3 suggest that the frequency of interaction affects how communicative signals develop; chapter 4 highlights how the modality of articulation can lead to specific effects on the structures that emerge. Though this insight is itself a contribution of the thesis, it does highlight that the work presented here is not the whole story, and further work can shed light on the particular interactions between the mechanisms I have investigated here.

The cultural evolutionary mechanisms that are the focus of this thesis, interaction and transmission, are complex social processes that have been rendered more simply here. The transmission fidelity of communicative signals between model and learner can vary, depending on who or how many people an individual learns from (Smith et al., 2016). The full scope of interactive behaviours accounts for more than just the exchange of verbal or manual signals, but also includes eye-gaze, body position and the negotiation of turn-taking (Hanna & Brennan, 2007; Healey et al., 2007; Stivers et al., 2009). As such, the present work does not fully capture these processes, but observes interaction and transmission in a smaller range of contexts than they appear in the real world. Future work (suggested in section 5.3) can apply artificial sign language learning experiments to further analyse different instantiations of interaction and transmission.

Finally, artificial sign language learning faces similar shortcomings to other experimental frameworks. The lab is not a natural environment, and may therefore lack some of the processes that occur in natural language creation. For example, these experiments use adult learners as participants, who can make use of existing linguistic knowledge. Although this method is designed to experimentally model processes of language emergence with the specific goal of reducing the interference from existing linguistic competence, it is not possible to completely remove the native language of participants in an experimental task. However, these experiments were not designed to be viewed only in isolation, but to be taken as a part of a wider body of research into language emergence that takes into account previous language learning experiments, as well as natural language data. In particular, these experiments are informed by, motivated by and designed to go hand in hand with data collected from naturally emerging sign languages, to provide a more complete view of the cultural evolution of language from its development in the individual learner to its evolution in populations of learners. A central goal of experimental research is to shed light where real-world complexities leave a problem obscured; the complementary analyses of experimental research and natural language research should provide solutions to each other's limitations, and thus provide a more complete resolution to the question of interest.

5.3 Future applications for artificial sign language learning

The flexibility of artificial sign language learning, as with other artificial language learning paradigms, lends itself well to expansion over a wide range of experimental contexts.

Future research can address some of the limitations I have discussed above, in section 5.2. For example, to further understand the roles that interaction and transmission play in the emergence of systematic structure, it is necessary to provide more experimental comparisons between situations where only interaction or only transmission is at play, in contrast to both processes in combination. Future studies could contrast learning from multiple models in comparison to only one training model, for instance, or could test the effect of learning from, and interacting with, communicators of different status (contrasting a parental relationship with the relationship of a peer). Importantly, social contexts should be contrasted across comparable experimental settings, to understand how each mechanism operates in the same linguistic context. In addition, evidence of the effects of these cultural evolutionary mechanisms should be contrasted with comparable experiments of language production and processing in the individual, in order to clarify the impact of individual cognition on language evolution, in comparison with population-level processes.

Artificial sign language learning also serves to elucidate the relationships between interaction and transmission, and how different instantiations of these processes affect linguistic output. For example, previous research, including research into emerging sign languages, has suggested that community structure can effect linguistic structure (Lupyan & Dale, 2010; Meir, Assaf et al., 2005; Meir, Sandler et al., 2005; A. Senghas, 2005; Wray & Grace, 2007). This can be modelled experimentally, varying the number of participants who interact with each other, to model the difference between large and small community sizes. Artificial sign language learning provides a paradigm in which such a model can be implemented, providing a comparison of community structures that complements emerging sign language research.

Similarly, there remain many unanswered questions about the impact of learning from a linguistic model under different conditions. For example, how does the status of the model (a parent, or a peer) affect the productions of the learner? How does the communicative context of an individual learner, or language user, affect the linguistic system? Research on homesign systems suggests that parental input has little impact on the structure found in those systems (Goldin-Meadow & Mylander, 1983). Village sign languages, a sub-category of emerging sign languages, have high proportions of hearing signers (Meir, Assaf et al., 2005), and little is understood about their impact on the languages they use. Studying different types of learners in different social contexts presents avenues of research well-suited to implementation with artificial sign language learning.

The capacity for visual interaction in artificial sign language learning experiments, through streaming between networked computers, also allows the investigation of the full scale of interactive behaviours. The minimal interaction paradigm explored here can be expanded to analyse non-linguistic or paralinguistic behaviours such as eye-gaze, repair and feedback strategies. Further, directed research could examine how these behaviours, which are frequently grounded in the communicator's physicality, can be incorporated into a linguistic system, and shed further light on the emergence of linguistic communication.

The research I have presented in this thesis has demonstrated the application of a novel methodology for investigating the cultural evolution of language. It offers a substantial expansion of previous experimental research, as well as a valuable companion framework to research on naturally emerging languages. The wide applicability of this method opens up new pathways for understanding in more detail the roles that interaction between language users and transmission of a language to new learners play in the emergence of systematic linguistic structure, offering insight into fundamental questions about the evolution of human language.

References

Abner, N., Flaherty, M., Stangl, K., Brentari, D. & Goldin-Meadow, S. (2015). Emergent lexical categories: Nouns and verbs in homesign and Nicaraguan Sign Language. In *Formal and experimental advances in sign language theory (feast).* Barcelona, Spain..

Abner, N., Flaherty, M., Stangl, K., Brentari, D. & Goldin-Meadow, S. (2016). Distinguising objects and actions in homesign and Nicaraguan Sign Language: the emergence of lexical categories. In *12th conference on theoretical issues in sign language research.* Melbourne, Australia.

Arbib, M. A., Liebal, K. & Pika, S. (2008). Primate Vocalization, Gesture, and the Evolution of Human Language. *Current Anthropology*, *49*(6), 1053–1076.

Aronoff, M., Meir, I., Padden, C. & Sandler, W. (2003). Classifier constructions and morphology in two sign languages. In K. Emmory (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 53–84). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Aronoff, M., Meir, I. & Sandler, W. (2005). The Paradox of Sign Language Morphology. *Language*, *81*(2), 301–344.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Baxter, G. J., Blythe, R. A., Croft, W. & McKane, A. J. (2006). Utterance selection model of language change. *Physical Review E*, *73*(4), 46118.

Bickerton, D. (1990). *Language and Species*. Chicago: University of Chicago Press.

Blythe, R. & Croft, W. (2016). S-curves and the mechanisms of propagation in language change. *Language*, *88*(2), 269–304.

Botha, R. (2007). On homesign systems as a potential window on language evolution. *Language & Communication*, *27*(1), 41–53.

Brennan, S. E. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology. Learning, memory, and cognition*, *22*(6), 1482–1493.

Brentari, D., Coppola, M., Cho, P. W. & Senghas, A. (2016). Handshape complexity as a precursor to phonology: Variation, Emergence, and Acquisition. *Language Acquisition*, 1-24. (Advance online publication.)

Brentari, D., Coppola, M., Jung, A. & Goldin-Meadow, S. (2013). Acquiring word class distinctions in American Sign Language: Evidence from handshape. *Language learning and development : the official journal of the Society for Language Development*(2), 130–150.

Bronowski, J. & Bellugi, U. (1970). Language, Name, and Concept. *Science*, *180*, 669–673.

Caldwell, C. A. & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PloS one*, *7*(8), e43807.

Carr, J. W., Smith, K., Cornish, H. & Kirby, S. (2016). The Cultural Evolution of Structured Languages in an Open-Ended, Continuous World. *Cognitive Science*, *12371*, 1–32.

Carrigan, E. M. & Coppola, M. (2017). Successful communication does not drive language development: Evidence from adult homesign. *Cognition*, *158*, 10–27.

Carstensen, A., Xu, J., Smith, C. T. & Regier, T. (2015). Language evolution in the lab tends toward informative communication. In D. Noelle et al. (Eds.), *Proceedings of the 37th annual meeting of the cognitive science society.* (pp. 303–308). Pasadena, California: Cognitive Science Society.

Christensen, P., Fusaroli, R. & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, *146*, 67–80.

Christiansen, M. H. & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 458–489.

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

Clay, Z., Pople, S., Hood, B. & Kita, S. (2014). Young Children Make Their Gestural Communication Systems More Language-Like: Segmentation and Linearization of Semantic Elements in Motion Events. *Psychological science*.

Coppola, M. & Newport, E. L. (2005). Grammatical subjects in homesign: abstract linguistic structure in adult primary gesture systems without linguistic input. *Proceedings of the National Academy of Sciences*, *102*(52), 19259–19253.

Corballis, M. C. (2003). *From hand to mouth: the origins of language*. Princeton: Princeton University Press.

Corballis, M. C. (2010). The gestural origins of language. *WIRES Cognitive Science*, *1*, 2–7.

Cormier, K., Fenlon, J. & Schembri, A. (2015). Indicating verbs in British Sign Language favour motivated use of space. *Open Linguistics*, *1*, 684–707.

Cornish, H., Smith, K. & Kirby, S. (2013). Systems from Sequences: an Iterated Learning Account of the Emergence of Systematic Structure in a Non-Linguistic Task. In M. Knauff, M. Pauen, N. Sebanz & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the cognitive science society* (pp. 340–345). Austin, Texas: Cognitive Science Society.

Croft, W. (2000). *Explaining language change: an evolutionary account*. London: Longman Linguistics Library.

Culbertson, J. & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, 1–6.

Culbertson, J. & Kirby, S. (2016). Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. *Frontiers in Psychology*, *6*(January), 1–11.

Culbertson, J., Smolensky, P. & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306–329.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*, 369–382.

de Vos, C. (2011). A Signers' Village in Bali, Indonesia. *Minpaku Anthropology Newsletter*, *33*, 4–5.

de Vos, C. (2014). The Kata Kolok Pointing System: Morphemization and Syntactic Integration. *Topics in cognitive science*, 1–19.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H. & Monaghan, P. (2015). Systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615.

Dryer, M. S. (2011). Order of subject, object and verb. In M. Haspelmath (Ed.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from `http://wals.info/chapter/81`

Emmorey, K., Tversky, B. & Taylor, H. A. (2000). Using space to describe space: Perspective in speech, sign, and gesture. *Spatial Cognition and Computation*, *2*(3), 157–180.

Evans, N. (2000). Word classes in the world's languages. In G. E. Booij, C. Lehmann & J. Mugdan (Eds.), *Morphologie: Ein internationales handbuch zur flexion und wortbildung / morphology: International handbook on inflection and word- formation* (pp. 708–732). Berlin: De Gruyter.

Evans, N. & Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behavioral and brain sciences*, *32*(5), 429–48.

Fay, N., Arbib, M. & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive science*, *37*(7), 1356–67.

Fay, N. & Ellison, T. M. (2013). The cultural evolution of human communication systems in different sized populations: usability trumps learnability. *PloS one*, *8*(8), e71781.

Fay, N., Garrod, S., Roberts, L. & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive science*, *34*(3), 351–86.

Fedzechkina, M., Jaeger, T. F. & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(44), 17897–902.

Fehér, O., Wonnacott, E. & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*, *91*, 158–180.

Fodor, J. A. & Pylyshyn, Z. W. (1988). Connectionism and cognitive architechture: A critical analysis. *Cognition*, *28*, 3–71.

Frishberg, N. (1975). Arbitrariness and Iconicity: Historical Change in American Sign Language. *Language*, *51*(3), 696–719.

Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E. & Gibson, E. (2014). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition*, *136C*, 215–221.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, *29*(5), 737–767.

Galantucci, B. & Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in human neuroscience*, *5*, 11.

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.

Gershkoff-Stowe, L. & Goldin-Meadow, S. (2002). Is there a natural order for expressing semantic relations? *Cognitive psychology*, *45*(3), 375–412.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E. & Saxe, R. (2013). A Noisy-Channel Account of Crosslinguistic Word-Order Variation. *Psychological Science*, *24*(7), 1079–1088.

Gil, D. (1994). The structure of Riau Indonesian. *Nordic Journal of Linguistics*, *17*(02), 179–200.

Gillespie-Lynch, K., Greenfield, P. M., Feng, Y., Savage-Rumbaugh, S. & Lyn, H. (2013). A cross-species study of gesture and its role in symbolic development: implications for the gestural theory of language evolution. *Frontiers in psycho-*

logy, *4*(June).

Givon, T. (1985). Iconicity, isomorphism and non-arbitrary coding in syntax. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 187–219). Amsterdam: John Benjamins Publishing Company.

Goldin-Meadow, S. (2003). *The Resilience of Language*. New York, NY: Taylor and Francis.

Goldin-Meadow, S. (2015). The Impact of Time on Predicate Forms in the Manual Modality: Signers, Homesigners, and Silent Gesturers. *Topics in Cognitive Science*, *7*, 169–184.

Goldin-Meadow, S., Brentari, D., Coppola, M., Horton, L. & Senghas, A. (2014). Watching language grow in the manual modality: Nominals, predicates, and handshapes. *Cognition*, *136*, 381–395.

Goldin-Meadow, S., Butcher, C., Mylander, C. & Dodge, M. (1994). Nouns and Verbs in a Self-Styled Gesture System: What's in a name? *Cognitive psychology*, *27*, 259–319.

Goldin-Meadow, S. & Feldman, H. (1977). The development of language-like communication without a language model. *Science*, *197*(4301), 401–403.

Goldin-Meadow, S. & Mylander, C. (1983). Gestural communication in deaf children: noneffect of parental input on language development. *Science*, *221*(4608), 372–374.

Goldin-Meadow, S. & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, *391*(6664), 279–281.

Goldin-Meadow, S., Mylander, C. & Butcher, C. (1995). The resilience of combinatorial structure at the word level: morphology in self-styled gesture systems. *Cognition*, *56*(3), 195–262.

Goldin-Meadow, S., So, W. C., Ozyürek, A. & Mylander, C. (2008). The natural order of events: how speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(27), 9163–8.

Groce, N. E. (1985). *Everyone here spoke sign language: hereditary deafness on Martha's Vineyard*. Cambridge, MA: Harvard University Press.

Hall, M. L., Mayberry, R. I. & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*(1), 1–17.

Hanna, J. E. & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615.

Haviland, J. B. (2013). The emerging grammar of nouns in a first generation sign language: Specification, iconicity, and syntax. *Gesture*, *13*, 309–353.

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.

Heine, B. & Kuteva, T. (2007). *The Genesis of Grammar: a reconstruction*. Oxford: Oxford University Press.

Hobaiter, C. & Byrne, R. W. (2014). The Meanings of Chimpanzee Gestures. *Current Biology*, 1–5.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*, 88–96.

Hockett, C. F. (1977). *The view from language: selected essays, 1948-1974*. Athens, GA: The University of Georgia Press.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88.

Hopper, P. J. & Thompson, S. A. (1985). The iconicity of the universal categories of 'noun' and 'verbs'. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 151–183). Amsterdam: John Benjamins Publishing Company.

Horton, L., Goldin-Meadow, S., Coppola, M., Senghas, A. & Brentari, D. (2015). Forging a morphological system out of two dimensions: Agentivity and number. *Open Linguistics*, *1*(1), 596–613.

Hudson Kam, C. L. (2015). The impact of conditioning variables on the acquisition of variation in adult and child learners. *Language*, *91*(4), 906–937.

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*(2), 151–195.

Hunsicker, D. & Goldin-Meadow, S. (2013). How handshape type can distinguish between nouns and verbs in homesign. *Gesture*, *13*(3), 354–376.

Hurford, J. R. (2007). *The origins of meaning: language in the light of evolution*. Oxford: Oxford University Press.

Hwang, S.-O., Tomita, N., Morgan, H., Ergin, R., Ilkbasaran, D., Seegers, S., ... Padden, C. (2016). Of the body and the hands: patterned iconicity for semantic categories. *Language and Cognition*, 1–30.

Johnson, K. (2004). On the systematicity of language and thought. *Journal of Philosophy*, *101*(3), 111–139.

Kegl, J., Senghas, A. & Coppola, M. (1999). Creation through contact: Sign language emergence and sign language change in Nicaragua. In M. DeGraff (Ed.), *Language creation and language change: Creolization, diachrony and development* (pp. 179–237). Cambridge, MA: MIT Press.

Kemp, C. & Regier, T. (2012). Kinship Categories Across Languages Reflect General Communicative Principles. *Science*, *336*(6084), 1049–1054.

Kendon, A. (2016). Reflections on the "gesture-first" hypothesis of language origins. *Psychonomic Bulletin & Review*(1960).

Kimmelman, V. (2009). Parts of speech in Russian Sign Language: The role of iconicity and economy. *Sign language and linguistics*, *12*(2), 161–186.

Kirby, S. (2016). *VideoBox.* Edinburgh: University of Edinburgh. Retrieved from `http://edin.ac/2haREUz`

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*(31), 10681–6.

Kirby, S., Griffiths, T. & Smith, K. (2014). Iterated learning and the evolution of language. *Current opinion in neurobiology*, *28C*, 108–114.

Kirby, S., Tamariz, M., Cornish, H. & Smith, K. (2015). Compression and Communication in the Cultural Evolution of Linguistic Structure linguistic structure. *Cognition*, *141*, 87–102.

Klima, E. S. & Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.

Kocab, A., Pyers, J. & Senghas, A. (2014). Referential Shift In Nicaraguan Sign Language: A Transition From Lexical To Spatial Devices. *Frontiers in Psychology*.

Kocab, A., Senghas, A. & Snedeker, J. (2016a). The emergence of temporal language in Nicaraguan Sign Language. *Cognition*, *156*, 147–163.

Kocab, A., Senghas, A. & Snedeker, J. (2016b). Recursion in Nicaraguan Sign Language. In D. Grodner, D. Mirman, A. Papafragou & J. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 1343–1348). Philadelphia, Pennsylvania: Cognitive Science Society.

Krauss, R. M. & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*(3), 343–346.

Langus, A. & Nespor, M. (2010). Cognitive systems struggling for word order. *Cognitive Psychology*, *60*(4), 291–318.

Liddell, S. K. (2003). *Grammar, gesture and meaning in American Sign Language*. Cambridge: Cambridge University Press.

Lillo-Martin, D. & Meier, R. P. (2011). On the linguistic status of 'agreement' in sign languages. *Theoretical Linguistics*, *37*(3-4), 95–141.

Lupyan, G. & Dale, R. (2010). Language Structure Is Partly Determined by Social

Structure. *PLoS ONE*, *5*(1), 1–10.

Marno, H., Langus, A., Omidbeigi, M., Asaadi, S., Seyed-Allaei, S. & Nespor, M. (2015). A new perspective on word order preferences: the availability of a lexicon triggers the use of SVO word order. *Frontiers in Psychology*, *6*(August), 1–8.

Mathur, G. & Rathmann, C. (2012). Verb agreement. In R. Pfau, M. Steinbach & B. Woll (Eds.), *Sign language: An international handbook* (pp. 136–157). Berlin: De Gruyter Mouton.

Matthews, R. J. (1997). Can Connectionists Explain Systematicity? *Mind and Language*, *12*(2), 154–177.

Maurits, L., Perfors, A. & Navarro, D. (2010). Why are some word orders more common than others? A uniform information density account. *Advances in Neural Information Processing Systems*, *23*, 1585–1593.

Mayberry, R. I. (2010). Early Language Acquisition and Adult Language Ability: What Sign Language Reveals About the Critical Period for Language. *The Oxford handbook of deaf studies, language, and education*, *2*, 281–291.

Mayberry, R. I., Lock, E. & Kazmi, H. (2002). Linguistic ability and early language exposure. *Nature*, *417*(6884), 38.

McNeill, D. (2000). Introduction. In D. McNeill (Ed.), *Language and gesture* (pp. 1–10). Cambridge: Cambridge University Press.

Meier, R. P. (2004). Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech. *Modality and structure in signed and spoken language*, 1–25.

Meir, I. (2012). The evolution of verb classes and verb agreement in sign languages. *Theoretical Linguistics*, *38*(1-2), 145–152.

Meir, I., Aronoff, M., Börstell, C., Hwang, S.-o., Ilkbasaran, D., Kastner, I., . . . Sandler, W. (2014). The effect of being human and the basis of grammatical word order: insights from novel communication systems and young sign languages. *Cognition*, *158*(April), 1–40.

Meir, I., Aronoff, M., Sandler, W. & Padden, C. (2010). Sign languages and compounding. *Cross-disciplinary issues in compounding*, 301–322.

Meir, I., Assaf, I., Sandler, W., Padden, C. & Aronoff, M. (2005). The influence of community on language structure. *Linguistic Variation*, *12*(2), 247–291.

Meir, I., Padden, C., Aronoff, M. & Sandler, W. (2007). Body as subject. *Journal of Linguistics*, *43*(03), 531–563.

Meir, I., Sandler, W., Padden, C. & Aronoff, M. (2005). Emerging Sign Languages. In *Oxford handbook of deaf studies, language and education* (Vol. 2, pp. 1–22).

Mesoudi, A. (2015). Cultural Evolution: A Review of Theory, Findings and Controversies. *Evolutionary Biology*.

Mesoudi, A. & Whiten, A. (2008). Review. The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical transactions of the Royal Society B*, *363*(1509), 3489–501.

Miles, H. L. W. (1990). The cognitive foundations for reference in a signing orangutan. In S. T. Parker & K. R. Gibson (Eds.), *"language" and intelligence in monkeys and apes: Comparative developmental perspectives* (pp. 511–539). New York, NY, US: Cambridge University Press.

Monaghan, P., Shillcock, R. C., Christiansen, M. H. & Kirby, S. (2014). How arbitrary is language? *Philosophical transactions of the Royal Society B*, *369*(1651), 20130299–.

Morford, J. P. & Goldin-Meadow, S. (1997). From here and now to there and then: the development of displaced reference in homesign and English. *Child development*, *68*(3), 420–35.

Morgan, G., Herman, R. & Woll, B. (2002). The development of complex verb constructions in British Sign Language. *Journal of child language*, *29*(3), 655–675.

Morgan, H. E. (2015). When Does a Word Emerge ? Lexical and Phonological Variation in a Young Sign Language. *San Diego Linguistic Papers*(5), 2–17.

Morgan, H. E. & Mayberry, R. I. (2012). Complexity in two-handed signs in Kenyan Sign Language: Evidence for sublexical structure in a young sign language. *Sign Language & Linguistics*, *15*(1), 147–174.

Motamedi, Y., Schouwstra, M. & Kirby, S. (in press). An evolutionary approach to sign language emergence: from state to process. [Peer commentary on "Gesture, sign and language: the coming of age of sign language and gesture studies"]. *Behavioral and Brain Sciences*.

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J. & Kirby, S. (under revision). Evolving artificial sign languages in the lab: from improvised gesture to systematic sign.

Namboodiripad, S., Lenzen, D., Lepic, R. & Verhoef, T. (2016). Measuring Conventionalization in the Manual. *Journal of Language Evolution*, *0*(0), 1–10.

Napoli, D. J. & Sutton-Spence, R. (2014). Order of the major constituents in sign languages: Implications for all language. *Frontiers in Psychology*, *5*(MAY), 1–18.

Nyst, V. (2010). Sign languages in West Africa. *Sign languages*, 405–432.

Özcalışkan, S., Lucero, C. & Goldin-Meadow, S. (2016). Does language shape silent gesture? *Cognition*, *148*, 10–18.

Özyürek, A., Furman, R. & Goldin-Meadow, S. (2014). On the way to language: event segmentation in homesign and gesture. *Journal of child language*(May 2016), 1–31.

Padden, C. (1986). Verbs and role-shifting in American Sign Language. In C. Padden (Ed.), *Proceedings of the fourth national symposium on sign language research and teaching* (pp. 44–57). Washington DC: The National Association for the Deaf.

Padden, C. (1990). The Relation Between Space and Grammar in ASL Verb Morphology. In C. Lucas (Ed.), *Proceedings of the 2nd international conference on theoretical issues in sign language research* (pp. 118–132). Washington DC: Gallaudet University Press.

Padden, C., Hwang, S.-O., Lepic, R. & Seegers, S. (2014). Tools for Language: Patterned Iconicity in Sign Language Nouns and Verbs. *Topics in cognitive science*, 1–14.

Padden, C., Meir, I., Aronoff, M. & Sandler, W. (2010). The grammar of space in two new sign languages. In D. Brentari (Ed.), *Sign languages: A cambridge language survey.* (pp. 570–592). Cambridge: Cambridge University Press.

Padden, C., Meir, I., Hwang, S.-O., Lepic, R., Seegers, S. & Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture*, *13*, 287–308.

Patterson, F. G. (1978). The gestures of a gorilla: Language acquisition in another pongid. *Brain and Language*, *5*(1), 72–97.

Peirce, J. W. (2007). PsychoPy–Psychophysics software in Python. *Journal of neuroscience methods*, *162*(1-2), 8–13.

Perfors, A. & Navarro, D. J. (2014). Language Evolution Can Be Shaped by the Structure of the World. *Cognitive Science*, *38*(4), 1–19.

Perlman, M. & Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*, *14*(3), 320–350. doi: 10.1075/gest.14.3.03per

Perniss, P., Thompson, R. L. & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*(Dec), 1–15.

Pichler, C. (2012). 28. Acquisition. *Sign Language: An International Handbook*.

Pinker, S. & Bloom, P. (2000). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*, 707–784.

Pullum, G. K. & Scholz, B. C. (2007). Systematicity and natural language syntax. *Croatian Journal of Philosophy*, *7*(21), 375–402.

R Core Development Team. (2008). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Ravignani, A., Delgado, T. & Kirby, S. (2016). Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behaviour*, *1*, 7.

Regier, T., Kemp, C. & Kay, P. (2015). Word meanings across languages support efficient communication. In B. McWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 237–263). Hoboken, NJ: Wiley-Blackwell.

Richie, R., Yang, C. & Coppola, M. (2014). Modeling the emergence of lexicons in homesign systems. *Topics in cognitive science*, *6*(1), 183–95.

Roberts, A., Vick, S.-J., Roberts, S. G. B. & Menzel, C. R. (2014). Chimpanzees modify intentional gestures to coordinate a search for hidden food. *Nature communications*, *5*, 3088.

Roberts, G. & Galantucci, B. (2012). The emergence of duality of patterning: Insights from the laboratory. *Language and Cognition*, *4*(4), 297–318.

Roberts, G., Lewandowski, J. & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, *141*, 52–66.

Saffran, J. R., Newport, E. L. & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621.

Sandler, W. (2013). Vive la différence: Sign language and spoken language in language evolution. *Language and Cognition*, *5*(2-3), 189–203.

Sandler, W., Aronoff, M., Meir, I. & Padden, C. (2011). The gradual emergence of phonological form in a new Language. *Natural Language and Linguistic Theory*, *29*(2), 503–543.

Sandler, W. & Lillo-Martin, D. C. (2006). *Sign language and linguistic universals*. New York, NY: Cambridge University Press.

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.

Schouwstra, M. (2012). *Semantic Structures, Communicative Strategies and the Emergence of Language*. Utrecht: LOT Dissertation Series. (no. 312.)

Schouwstra, M. (2016). Temporal structure in emerging language : from natural data to silent gesture. *Cognitive science*.

Schouwstra, M. & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–6.

Schouwstra, M., Smith, K. & Kirby, S. (2016). From natural order to convention in silent gesture. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barcelo-Coblijn, O. Feher & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 11th international conference (evolang11)*. Retrieved from `http://evolang.org/neworleans/papers/67.html`

Scott-Phillips, T. C., Kirby, S. & Ritchie, G. R. S. (2009). Signalling signalhood and the emergence of communication. *Cognition*, *113*(2), 226–233.

Senghas, A. (2000). The Development of Early Spatial Morphology in Nicaraguan Sign Language. In S. C. Howell, S. A. Fish & T. Keith-Lucas (Eds.), *Proceedings of the boston university conference on language development* (p. 696-707). Boston: Cascadilla Press.

Senghas, A. (2005). Language Emergence: Clues from a New Bedouin Sign Language. *Current Biology*, *15*(12), 463–465.

Senghas, A. & Coppola, M. (2001). Children creating language: how Nicaraguan sign language acquired a spatial grammar. *Psychological science*, *12*(4), 323–8.

Senghas, A., Coppola, M., Newport, E. L. & Supalla, T. (1997). Argument structure in Nicaraguan Sign Language: the emergence of grammatical devices. In E. Hughes, M. Hughes & A. Greenhill (Eds.), *Proceedings of the boston university conference on language development* (pp. 550–561). Boston: Cascadilla Press.

Senghas, A., Kita, S. & Ozyürek, A. (2004). Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science*, *305*(5691), 1779–82.

Senghas, A., Ozyürek, A. & Goldin-Meadow, S. (2013). Homesign as a way station between co-speech gesture and sign language: the evolution of segmentation and sequencing. In R. Botha & M. Everaert (Eds.), *The evolutionary emergence of language: evidence and influence* (pp. 62–76). Oxford: Oxford University Press.

Senghas, R. J., Senghas, A. & Pyers, J. E. (2005). The Emergence of Nicaraguan Sign Language: Questions of Development, Acquisition, and Evolution. *Biology and knowledge revisited: From neurogenesis to psychogenesis*(April), 287–306.

Siewierska, A. (2013). Verbal Person Marking. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from `http://wals.info/chapter/102`

Silvey, C., Kirby, S. & Smith, K. (2014). Word Meanings Evolve to Selectively Preserve Distinctions on Salient Dimensions. *Cognitive science*, 1–15.

Singleton, J. L., Goldin-Meadow, S. & McNeill, D. (1995). The cataclysmic break between gesticulation and sign: evidence against a unified continuum of gestural communication. In K. Emmorey & J. Reilly (Eds.), *Language, gesture and space* (pp. 287–311). Hillsdale, NJ: Erlbaum Associates.

Singleton, J. L. & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, *49*(4), 370–407.

Slobin, D. I. (2004). From ontogenesis to phylogenesis: what can child language tell us about language evolution? In J. Langer, S. Parker & C. Milbrath (Eds.), *Biology and knowledge revisited: from neurogenesis to psychogenesis* (pp. 255–285). Lawrence Erlbaum Associates.

Slocombe, K. (2012). Have we underestimated great ape vocal capacities. In K. R. Gibson & M. Tallerman (Eds.), *Oxford handbook of language evolution.* (pp. 90–95). Oxford: Oxford University Press.

Smith, K., Abramova, K., Cartmill, E. & Kirby, S. (under revision). The experimental study of sign language emergence.

Smith, K., Kirby, S. & Brighton, H. (2003). Iterated learning: a framework for the emergence of language. *Artificial life*, *9*, 371–386.

Smith, K., Perfors, A., Feher, O., Samara, A., Swoboda, K. & Wonnacott, E. (2016). Language learning , language use , and the evolution of linguistic variation. *Philosophical transactions of the Royal Society B*, 1–16.

Smith, K. & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–9.

So, W. C., Coppola, M., Licciardello, V. & Goldin-Meadow, S. (2005). The seeds of spatial grammar in the manual modality. *Cognitive science*, *29*(6), 1029–43.

Spaepen, E., Coppola, M., Flaherty, M., Spelke, E. & Goldin-Meadow, S. (2013). Generating a lexicon without a language model: Do words for number count? *Journal of memory and language*, *69*(4), 496–505.

St. Clair, M. C., Monaghan, P. & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science*, *33*(7), 1317–1329.

Stadler, K., Blythe, R. A., Smith, K. & Kirby, S. (2016). Momentum in Language Change. *Language Dynamics and Change*, *6*(2), 171–198.

Sterelny, K. (2012). Language, gesture, skill: the co-evolutionary foundations of language. *Philosophical transactions of the Royal Society B*, *367*(1599), 2141–51.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10587–92.

Stokoe, W. C. (1960). *Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf.* (Vol. 8). Buffalo, NY: Department of Anthropology and Linguistics: University of Buffalo.

Supalla, T. & Newport, E. L. (1978). How many seats in a chair? The derivation of nouns and verbs in American Sign Language. In P. Siple (Ed.), *Understanding language through sign language research* (pp. 91–132). Rochester, NY: Academic

Press.

Sutton-Spence, R. & Woll, B. (1999). *The Linguistics of British Sign Language*. Cambridge: Cambridge University Press.

Tamariz, M. & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, *8*, 37–43.

Taub, S. (2001). *Language from the body: Iconicity and Metaphor in American Sign Language*. Cambridge: Cambridge University Press.

Theisen, C. A., Oberlander, J. & Kirby, S. (2010). Systematicity and arbitrariness in novel communication systems. *Interaction Studies*, *11*(1), 14–32.

Theisen-White, C., Kirby, S. & Oberlander, J. (2011). Integrating the horizontal and vertical cultural transmission of novel communication systems. In L. Carlson, C. Holscher & T. F. Shipley (Eds.), *Expanding the space of cognitive science: Proceedings of the 33rd annual meeting of the cognitive science society* (pp. 956–961). Boston, Massachusetts: Cognitive Science Society.

Thompson, B., Kirby, S. & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*(16), 201523631.

Tkachman, O. & Sandler, W. (2013). The noun–verb distinction in two young sign languages. *Gesture*, *13*(3), 253–286.

Verhoef, T., Kirby, S. & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, *43*, 57–68.

Verhoef, T., Kirby, S. & de Boer, B. (2015). Iconicity and the Emergence of Combinatorial Structure in Language. *Cognitive Science*, *40*, 1969–1994.

Winters, J., Kirby, S. & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, *7*(3), 1–35.

Wray, A. & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, *117*(3), 543–578.

Xu, J., Dowman, M. & Griffiths, T. L. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society B*, *280*(March), 20123073.

Zeshan, U. (2007). The ethics of documenting sign languages in village communities Language Documentation & Linguistic Theory. In P. K. Austin, O. Bond & D. Nathan (Eds.), *Proceedings of the conference on language documentation and linguistic theory* (pp. 269–279). London: SOAS.

Appendix A

Stimulus sentences from chapter 4

Below is the full set of sentences used as stimuli for the experiment described in chapter 4. Participants were presented with these sentences, in pairs, and asked to communicate them using only gesture.

Sentences pairs were organised into blocks of four, with each set containing a pair expressing each of four verb types: plain verbs of motion, verbs of motion with a recipient or goal, verbs of physical transfer, and verbs of non-physical transfer. A set of four sentences could be of one or two-types: same-agent (where the agent was the same in each sentence in a pair), or different-agent (where agents in each sentence in a pair were different. Sentences in the experiment were pre-defined, and the same set (those given below) were used for all participants in the experiment, though the order in which sets, pairs, and individual sentences were presented was randomised.

Set 1 (different-agent)

Plain verbs

Hannah is running
Sarah is cycling

Locative verbs

Hannah is walking to Sarah
Sarah is swimming to Hannah

Physical transfer verbs

Hannah is throwing a hat to Sarah
Sarah kicking a ball to Hannah

Non-physical transfer verbs

Hannah is scolding Sarah
Sarah is helping Hannah

Set 2 (different-agent)

Plain verbs

Hannah is swimming
Sarah is walking

Locative verbs

Hannah is cycling to Sarah
Sarah is running to Hannah

Physical transfer verbs

Hannah is giving a book to Sarah
Sarah writing a letter to Hannah

Non-physical transfer verbs

Hannah is phoning Sarah
Sarah is praising Hannah

Set 3 (same-agent, Hannah only)

Plain verbs

Hannah is cycling
Hannah is swimming

Locative verbs

Hannah is running to Sarah
Hannah is walking to Sarah

Physical transfer verbs

Hannah is kicking a ball to Sarah
Hannah is sending a letter to Sarah

Non-physical transfer verbs

Hannah is phoning Sarah
Hannah is scolding Sarah

Set 4 (same-agent, Sarah only)

Plain verbs

Sarah is walking
Sarah is running

Locative verbs

Sarah is swimming to Hannah
Sarah is cycling to Hannah

Physical transfer verbs

Sarah is giving a book to Hannah
Sarah is throwing a hat to Hannah

Non-physical transfer verbs

Sarah is praising Hannah
Sarah is helping Hannah

Appendix B

Guide to example videos

Videos can be accessed via the University of Edinburgh's DataShare facility, at `http://hdl.handle.net/10283/2338`**, or with the DOI** `http://dx.doi.org/10.7488/ds/1693`**.**

Videos depicting gestures can be accessed through the Edinburgh University DataShare repository. Video examples correspond to stills given in figures for each chapter, and details about how to identify video examples are given in each section below.

B.1 Examples from Chapter 2: Evolving artificial sign languages in the lab

Video file names

All video files relating to the experiments detailed in chapter 2 are prefixed with the tag **EX02**.
File names are given with a tag identifying the meaning of the gesture, the experimental condition, the chain or pair, and the round or generation. For the transmission plus interaction and interaction-only conditions, a further identifier marks the participant (simply as A or B). Seed gestures are marked with the meaning tag plus a 'SEED' tag (i.e. 'arrest_SEED'). For example, the file name

EX02_prison_IT_35A

refers to a gesture for the meaning prison, taken from a participant in the transmission plus interaction condition, from generation 5 of chain 3. File names are given in text to identify each video.

B.1.1 Seed gesture examples

Seed gestures were collected from 48 individuals, who produced a gesture for a single meaning in the meaning space, giving 2 gesture videos per meaning. Unique sets of 24 seed videos were produced for each chain by randomly selecting 1 out of the 2 gesture videos for each meaning in the meaning space.

Example of a seed gesture for *to make an arrest*

See figure 2.4.

File name:

EX02_arrest_SEED

The participant gestures pointing a gun at a perpetrator, running after the perpetrator and seizing and handcuffing said perpetrator.

Examples of seed gestures for the theme of hairdressing: *hairdresser*, *hair salon*, and *to give a haircut*

See figure 2.5.

File names:

EX02_hairdresser_SEED

EX02_hair_salon_SEED

EX02_haircut_SEED

All participants produce a repeated hair-cutting gesture that does not provide information to distinguish between the meanings.

B.1.2 Examples from experiment 1: transmission plus interaction

Examples for the meaning prison from chain 3, generations 0, 1 and 5.

See figure 2.6.

File names:

EX02_prison_SEED

EX02_prison_IT_31A

EX02_prison_IT_31B

EX02_prison_IT_35A

EX02_prison_IT_35B

The participant at generation 0 produces a highly iconic, pantomimic gesture. Gestures at generation retain a lot of the pantomimic qualities of the gesture, but reduce the number of elements. By generation 5, it has become a two part gesture, signalling the dimensions of the meaning space.

Examples showing the systematic reuse of signals on both the functional category and the thematic category.

See figure 2.7.

File names:

EX02_prison_IT_35B

EX02_church_IT_35B

EX02_hair_salon_IT_35B

EX02_photographer_IT_45B

EX02_camera_IT_45B

EX02_take_photo_IT_45B

For gestures in the functional category of *location*, a roof gesture is systematically re-used across the category.

For gestures in the thematic category of *photography*, the participant reuses the same camera gesture.

Examples showing the reanalysis of a wave gesture to a functional element, through a single chain.

See figure 2.8.

File names:

EX02_hairdresser_IT_21A

EX02_hairdresser_IT_22A

EX02_haircut_IT_25A

EX02_haircut_IT_25B

EX02_sing_IT_25A

EX02_sing_IT_25B

A waving gesture is mimed in generation 1 for the meaning *hairdresser*, in which the participant mimes being a hairdresser and waving at a client. By generation 5, the gesture has changed formally and been reanalysed as a functional marker, evidenced by its use in a gesture sequence for the meaning *to sing*.

B.1.3 Examples from the interaction-only condition (experiment 2)

Examples showing idiosyncratic gestures without the re-use of functional elements

See figure 2.14.

File names:

EX02_prison_I_31A

EX02_prison_I_35A

EX02_church_I_35A

Gestures become reduced in form between rounds 1 and 5, and the participant does not signal the similarity in meaning between *prison* and *church*.

Examples showing reduction in form of gestures

See figure 2.15.

File names:

EX02_photographer_I_11A

EX02_photographer_I_15A

EX02_camera_I_15A

At round 1, the participant uses an iconic point-at-self gesture, followed by a camera outline, to convey *photographer*. By generation 5, the camera outline has been reduced, showing greater ambiguity. She does not re-use thematic elements across meanings, as shown by her gesture for *camera*.

B.1.4 Examples from the transmission-only condition (experiment 2)

Example of gestures in the *location category*, at generation 1 and 5 of the same chain.

See figure 2.12.

File names:

EX02_prison_T_41

EX02_prison_T_45

EX02_church_T_45

The gesture for *prison* at generation 1, like in other conditions, is still highly iconic and pantomimic. By generation 5, the gesture has become segmented, re-using and re-combining elements. The participant frequently repeats both thematic and function sub-gestures.

Examples of redundancy in gestures in the transmission-only condition

See figure 2.13.

File name:

EX02_photographer_T_55

In this gesture for *photographer*, the participant shows frequent re-use of sub-gestures, demonstrating redundancy in the gesture sequence.

B.2 Examples from Chapter 3: The emergence of linguistic categories

B.2.1 Video file names

Video file names for showing examples from chapter 3 are prefixed with **EX03**.

File names are given a meaning identify that identifies the target item and whether it is used in a typical or atypical event. The numeric identifier gives a unique number to each pair, with the addition of 1, 2 or 3 to denote each stage of the experiment (IP1, COMM or IP2), and a participant identifier (A or B). For example, the file name

EX03_camera_typical_31A

presents a gesture for the target item camera used in a typical scene. The gesture was produced by a participant in the stage IP1.

B.2.2 Examples showing the number of gestures denoting target elements

Examples of gestures where the target item is included in the gesture, and where the target item is omitted from the gesture

See figure 3.4.

File names:

EX03_camera_typical_51A

EX03_camera_atypical_51A

For the typical-use gesture, the participant produces a camera gesture, repeated. For the atypical-use gesture, the participant does not specify the target item of *camera*.

Examples of a gesture sequence where the participant produces two unique gestures for the target element

See figure 3.5.

File name:

EX03_sewingmachine_typical_212A

The participant produces two sub-gestures for the target element, one in which she depicts sewing with a needle and thread, and one in which she represents the motion of the sewing machine.

B.2.3 Examples of handshape strategies used to describe different scenes

Examples of handling handshapes used across scene types.

See figure 3.9a.

File names:

EX03_nailpolish_typical_112A

EX03_nailpolish_atypical_112A

For the both the typical-use and atypical-use scenes, the participant produces a handling gesture for the target item *nail polish*.

Examples of different handshapes used in different scene types.

See figure 3.9b.

File names:

EX03_machete_typical_92A

EX03_machete_atypical_92A

The participant produces a handling handshape for the typical-use scene *cut with machete*, but uses a descriptor gesture for the atypical-use scene *drop machete in bin*.

B.2.4 Example of the use of a basehand gesture.

See figure 3.10a.

File name:

EX03_hammer_typical_122A

The participant gestures the meaning *hammer against a wall*, using a basehand to ground the hammering gesture.

B.2.5 Examples of repetitions in gestures

Example of an iconic repetition

See figure 3.14a.

File name:

EX03_scissors_typical_202A

The participant produces a repeated cutting gesture, iconically represent the repetition inherent in cutting across a sheet of paper.

Example of a non-iconic repetition

See figure 3.14b.

File name:

EX03_ring_typical_162B

The participant repeats a gesture in which he puts on a ring, using a repeated action not represented in the target scene.

B.3 Examples from Chapter 4: The cultural evolution of complex constructions

B.3.1 Video file names

Video file names showing examples from chapter 4 are prefixed with **EX04**.
Video files are further identified by a set of strings that identifies the sentence pairs, with each sentence denoted by the agent of the sentence and a short code noting the verb (e.g. *Sthrow*). The participant identifier is given as the chain, the generation and an individuating label (A or B). For example, the file name

EX04_Scyc_Hrun_53A

denotes a gesture video for the target pair:

Sarah cycles to Hannah.
Hannah runs to Sarah.

The participant in this example took part in chain 5, generation 3 of the experiment.

B.3.2 Examples of gestures using the lexical strategy.

Examples of the use of 1- and 2-handshapes

See figure 4.7.

File name:

EX04_Scyc_H_run_53A

The participant uses the 1 handshape to refer to Sarah and the 2 handshape to denote Hannah, within a sentence pair.

Examples of lexical handshapes to denote specific agents.

See figure 4.8.

File names:

EX04_Hkick_Hsend_15B

EX04_Hthrow_Skick_55A

The first participant uses the 1-handshape to consistently denote *Hannah* in both sentences of a same-agent sentence pair. The second participant uses the 1-handshape to refer to the agent, *Hannah*, and the 2-handshape to denote the goal, *Sarah*.

B.3.3 Examples of gestures using the body strategy.

Examples of the use of body orientation to denote agents in a target sentence.

See figure 4.9.

File names:

EX04_Ssend_Hgive_25A

EX04_Hthrow_Skick_25A

The participant uses body orientation to denote different agents, orienting to the right for *Sarah*, and to the left for *Hannah*.

Examples of the body strategy in same-agent sentence pairs.

See figure 4.10

File names:

EX04_Hscold_Hphone_45B

EX04_Hscold_Hphone_25A

The first participant contrasts body orientation to denote the agent in a same-agent sentence pair (*Hannah*). The second participant does not use body orientation to distinguish between agents in a same-agent sentence pair, though she does so for different-agent sentence pairs.

B.3.4 Examples of gestures using the indexing strategy

Examples of the index strategy

See figure 4.11

File names:

EX04_Shelp_Hscold_32A

EX04_Hscold_Shelp_33A

The first participant uses an indexing stragy along the z-axis, point to herself for the agent and away from the body for the goal. The second participant indexes sentence participants along the x-axis.

Examples of consistent indexing locations

See figure 4.12

File names:

EX04_Sswim_Hwalk_34A

EX04_Hphone_Spraise_34A

The participant uses consistent locations for *Hannah* and *Sarah*, across sentence pairs.

Examples of verb movement with respect to index locations

See figure 4.13

File names:

EX04_Skick_Hthrow_32A

EX04_Hgive_Ssend_34B

The first participant produces a verb path which is neutral in relation to the index locations for *Hannah* and *Sarah*. The second participant produces a verb path that demonstrates movement between the indexed locations for *Hannah* and *Sarah*.