It's not all in your Mind:\\

Modality Effects on the Emergence of Combinatorial Structure

HANNAH LITTLE

Promoter: Prof. Dr. Bart de Boer

VUB 2017

It's not all in your Mind:
Modality Effects
on the Emergence of Combinatorial Structure

Hannah Little

Doctor of Sciences

Artificial Intelligence Laboratory, Department of Computer Science
Vrije Universiteit Brussel

2017

Abstract

Previous work in evolutionary linguistics has focused on cognitive biases in interaction with learning and communication. In contrast, work on the evolution of speech has focused on physical features of the vocal tract. In this thesis, I aim to bridge the gap between these two fields and investigate how a linguistic modality (speech or sign) affects the emergence of combinatorial structure in language. Using artificial language experiments with continuous audio signals, I manipulate signal spaces to investigate how features of linguistic modalities affect the emergence of combinatorial structure. In Experiment 1, the size of signal spaces is investigated. I show that in a smaller signal space, combinatorial structure emerges more reliably than in bigger signal spaces. Experiments 2 and 3 investigate the dimensionality of a signal space, finding more structural strategies when meaning dimensions outnumber signal dimensions. Experiment 4 demonstrates the effects of different types of iconicity on structure in signals, demonstrating that signals had more movement when signal-meaning mappings were less intuitive. Experiments 5 and 6 compare individual signal production and communication respectively in the process of conventionalisation in speech-like signals. These results are compared to existing experimental results from work on conventionalisation in which graphical signals lose their complexity and iconicity as the result of communication (Garrod et al., 2007). In my experiments, however, signals failed to become less complex both with and without communication, but overall, signals produced with communication were more complex. Experiment 7 measures iconicity in the signals produced in Experiments 5 and 6 by asking naïve participants to match signals to their meanings. Iconicity increased over time with communication, again suggesting signals were not being conventionalised. The incongruity between Garrod's (2007) results and my experiments may be due to the signals in my experiments being less flexible and iconic in the first instance, indicating effects of signalling modality in processes such as conventionalisation. Finally, Experiment 8 tests the effects of signal duration on structure in signals and shows that with a time constraint, signals had less complexity than those produced without. Overall, the experiments demonstrate that the features of a signalling modality affect signal-meaning mappings, having knock-on effects on signal structure. This has implications not only for the emergence of structure in real world languages, but also in the design of artificial signal spaces for experimental work, and the validity of generalisations from previous experimental results.

Lay Summary

Where did structure in language come from? Previous work in evolutionary linguistics has tackled this question focusing on cognitive biases in interaction with learning and communication. In contrast, work on the evolution of speech has focused on physical features of the vocal tract. In this thesis, I bridge the gap between these two fields and investigate how a linguistic modality (speech or sign) affects the emergence of combinatorial structure in language. This is the level of structure where meaningless building blocks combine to make meaningful utterances.

There have been some differences observed in how structure emerges and is used in different modalities in the real world. For example, in emerging sign languages a level of combinatorial structure is very slow to emerge. Here, I test different hypotheses about what may cause these differences.

Using artificial language experiments with continuous audio signals, I manipulate signal spaces to investigate how features of linguistic modalities affect the emergence of combinatorial structure. I investigate the effects of size and dimensionality of the signalling space, and how easy it is to generate clear signal-meaning mappings. I also consider how pressures for brevity, discrimination and processes such as conventionalisation affect the emergence of combinatorial structure.

Overall, the experiments demonstrate that the features of a signalling modality affect signal-meaning mappings, having knock-on effects on signal structure. This has implications not only for the emergence of structure in real world languages, but also in the design of artificial signal spaces for experimental work, and the validity of generalisations from previous experimental results.

Lekensamenvatting

Waar komt de structuur van taal vandaan? Aan de ene kant heeft eerder werk over de evolutie van grammatica deze vraag onderzocht door te kijken naar eigenschappen van cognitie en de interactie hiervan met het leren van taal en de functie van taal in communicatie. Aan de andere kant heeft onderzoek naar de evolutie van spraak gekeken naar het effect van de fysische eigenschappen van het spraakkanaal. In dit proefschrift probeer ik de kloof tussen deze twee benaderingen te overbruggen door te onderzoeken hoe de modaliteit van taal (spraak of visuele gebaren) het ontstaan van combinatorische structuur beïnvloedt. Combinatorische structuur is de eigenschap van taal dat betekenisloze bouwsteentjes worden gecombineerd to betekenisvolle uitingen.

Er zijn al een aantal verschillen bekend in het ontstaan en het gebruik van structuur in verschillende modaliteiten in menselijke taal. In gebarentalen die nieuw ontstaan ontwikkelt combinatorische structuur zich bijvoorbeeld maar langzaam. In dit proefschrift onderzoek ik verschillende hypothesen die deze verschillen kunnen verklaren.

In experimenten waarin proefpersonen kunstmatige taaltjes leren die bestaan uit continue signalen, manipuleer ik de mogelijke signalen die proefpersonen kunnen maken. Op die manier kan ik onderzoeken hoe de modaliteit (de aard van de signalen) invloed heeft op het ontstaan van combinatorische structuur. Met name onderzoek ik het effect van de grootte van de signaalruimte, hoeveel vrijheidsgraden die hebben en hoe moeilijk het is om een rechtstreeks verband te hebben tussen signalen en hun betekenissen. Ik onderzoek ook hoe druk voor bondigheid en onderscheidbaarheid invloed hebben, en hoe het proces van conventionalisering effect heeft op het ontstaan van combinatorische structuur.

De experimenten laten zien dat de aard van de signaalruimte duidelijke invloed heeft op de het verband tussen signalen en hun betekenissen, en daarmee ook het ontstaan van structuur beïnvloeden. Deze bevindingen zijn niet alleen relevant voor het begrijpen van het ontstaan van de structuur van echte menselijke talen, maar zijn ook van belang voor het ontwerpen van nieuwe experimenten met kunstmatige signalen en voor het evalueren van de conclusies die getrokken kunnen worden uit eerdere experimentele resultaten.

Acknowledgements

Thank you to the European Research Council for funding Bart de Boer's awesome ABACUS project, and thank you to Bart for using some of that funding to employ me. Thanks also to Bart for his supervision, for always having an open door, and for always responding to emails, questions and concerns with speed, care, intelligence, and never ending enthusiasm.

Thank you to the other members of the ABACUS project. I will miss beer days and never ending lunch conversations with Sabine van der Ham, Heikki Rasilo, Bill Thompson, Piera Filippi and Andrea Ravignani. A very special thanks go to Kerem Eryılmaz, without whom this project would have fallen on its arse long ago. I would like to thank Kerem for the experiment interface for the Leap Motion experiments, for helping me code, and for his assurance that I should give myself more credit when I did manage to code and fix stuff on my own.

The sheer number of people I have exchanged ideas with in the evolutionary linguistics community, at the EvoLang conferences and elsewhere, is ridiculous. I'd like to thank them all, but it would get silly fast, and I'd forget people and feel terrible, so here's a few special mentions. Thank you to Tessa Verhoef for the code for the slide whistle experiment interface. Thank you to Bodo Winter for help with statistics. Thank you to Yannick Jadoul for designing the graph in Figure 6.5. Thanks to Simon Kirby, Gareth Roberts and Monica Tamariz for specific ideas that came to fruition in this thesis. Thanks to Kenny Smith, the members of the LEC in Edinburgh, both past and present, and everyone who did the ELC MSc 2010-2011. You incited my love for language evolution in the first place, and have continued to throw coal on the fire ever since. Thanks to James Winters, Seán Roberts, Michael Pleyer and Stefan Hartmann, and everyone who has read, written or commented at *replicatedtypo.com*; our little blog has been the best source of ideas, discussions, and friendships. Thank you to all of my new friends at the Max Plank Institute in Nijmegen in the LEvInSoN group and elsewhere: Ashley Micklos, Yasamin Motamedi, Alan Nielsen, Marcus Perlman, Kevin Stadler, Justin Sulik, Mark Dingemanse, and Stephen Levinson himself! You have all been so welcoming, enthusiastic and brilliant.

A special thank you to Catriona Silvey and Ariana "Barry" Olsen for always being there, and for their patience with me continually talking about bees, politics, and language evolution, as well as for emotional support, draft reading, book recommendations, and nice times™. If my PhD was Lord of the Rings (it was), and I was Frodo

(I am), you would collectively be my Samwise. You couldn't write it for me, but you did carry me a lot of the way.

Thanks to all my linguist friends scattered across the world who still brighten my day when I hear from them; Jack Wilson, Christos Christodoulopoulos, Tom Fitz-Hugh, Robin Lindop Fisher, Daniel Schmidtke, countless linguist "tweeps", and anyone else who's ever engaged me in a conversation about language. My non-linguist, but no less awesome and supportive friends also deserve a mention; Tanya Kosáros, Theresia Feldmann, Ben Wheatley, Dan Etheridge, Drew Lott, Megan Sadler, Zeenat Rahim, Lydia Cope, Elise le Gat, Andrew Simmons and Mathew Ryalls, and everyone else who's ever let me crash on their sofa, or crashed on mine.

Thank you to the teachers back in Sixth Form who were the first to ever get me excited about linguistics: Harindar Uppal and Pamela Camm. And thank you also to all of my lecturers in York, Edinburgh, Berlin, Antwerp and Brussels, who kept my excitement going.

Thanks to all of my family for supporting my decision to leave the island, and for never understanding what I actually do, but never questioning that I should be doing it. You all mean so much, but again I will keep it to a few special mentions. My big sister, Emily Weight, has been a particular crutch for me when I've got myself in a state, and her and her husband Chris have never failed to make me feel at home. My twin, Rachel Little, is, and has always been, the winner, every birthday I spend away from her feels, and will always feel, wrong. My Dad, Rob Little, is a pillar of love, acceptance and support. My Mam, Ruth Nichol, has never ceased in showing her love, even when I've been at my most difficult. My Grandyma and Grandydad, Shirley and Ray Townsend, were a special presence in my intellectual growth as a child, especially regarding my love of language, books and libraries.

This final spot is where people usually thank their partner, but my love life has always been more of a hindrance to my productivity and mental health than a help. So in a somewhat unconventional move, I'd like my final thank you to be my counsellor in Brussels, Sanne. We should all talk about our feelings more, and I couldn't have done a lot of what I've done in the past few years without having an outlet of amazing, caring mental health professionals. Thank you to the Belgian government and the VUB for having this stuff available, accessible and free. It should always be available, accessible and free.

Take a piano. The keys begin, the keys end. You know there are 88 of them and no-one can tell you differently. They are not infinite, you are infinite. And on those 88 keys the music that you can make is infinite. I like that. That I can live by. But you get me up on that gangway and roll out a keyboard with millions of keys, and that's the truth, there's no end to them, that keyboard is infinite. But if that keyboard is infinite, there's no music you can play. You're sitting on the wrong bench. That's God's piano.

Danny Boodman T.D. Lemon 1900
from *The Legend of 1900* by Giuseppe Tornator
adapted from *Novecento, Un monologo* by Alessandro Baricco

Contents

CHAPTER 1

Introduction

Languages are made up of a finite number of meaningless building blocks. These building blocks can be combined in different ways to make meaningful items. For example, the phonemes /f/, /l/ and /iː/ can combine in English to make the words /fliː/ (flee), /liːf/ (leaf) or /fiːl/ (feel). This level of structure is called *combinatorial structure*. These words can then be combined to make the bigger meaningful sentences, such as: "The flee can feel the leaf". This level of structure is called *compositional structure*. Together, these two levels of structure make up the linguistic design feature *duality of patterning*, which is said to make language infinitely productive while using a very small finite set of meaningless units (Hockett, 1960). While this feature of language can be argued to be unique to language among other communication systems, Studdert-Kennedy & Goldstein (2003) argue that this combinatoriality comes from general principles of natural systems which are infinitely productive with finite means.

Where structure in language comes from is the focus of the field of *evolutionary linguistics* or *language evolution*. As there is very little direct evidence we can rely on from when language first emerged, much of this work uses indirect methods such as agent-based computational modelling (see Kirby, 2002, for a review) and experimental work with human participants learning or creating artificial languages (see Scott-Phillips & Kirby, 2010, for a review). The majority of this work, however, starts with a set of discrete units (usually letters) from which signals can be built in order to investigate the emergence of compositional structure. If we are instead interested in the emergence of combinatorial structure, defined by being the combination of meaningless building blocks, we cannot start with a set of meaningless building blocks. To tackle this problem, a growing body of both computational and experimental work is using continuous signal spaces (e.g. Verhoef et al., 2014). The use of continuous signal spaces allows us to model the process of categorisation of a signalling space and the emergence of combinatorial structure.

Previous work in the domain of language evolution has mainly focused on cognitive pressures, and pressures for expressivity and learning. However, there is also a substantial body of work on the evolution of speech, which focuses on the physical features of the vocal tract (see Fitch, 2000; de Boer, 2016 for discussion on why the evolution of speech and language should be investigated separately). Work on the evolution of speech is mostly preoccupied with the question of whether the human vocal tract evolved specifically for language. For example, did the human larynx descend to allow for more flexibility in our speech, or was it caused by some other evolutionary pressure? A question which is less discussed in both the evolution of language and speech is, regardless of whether the human vocal tract evolved specifically for language or not, did spoken language culturally evolve to adapt to the vocal tract and the oral articulators? The answer is, of course, yes. However, the specifics of how the vocal tract affects the structure we see in language, especially at a combinatorial level, and how this compares to the effects of other modalities, is largely unexplored territory for artificial language experiments.

Exploring the effects of the vocal tract on combinatorial structure in speech, or the effects of the sign modality, is difficult using artificial language experiments. This is because using existing language modalities is not methodologically ideal, as the pre-existing linguistic knowledge of human participants will interfere with their behaviour. Previously, researchers have looked at the differences between the linguistic structures produced using different language modalities (e.g. Meier, 2002). However, Galantucci et al. (2010) demonstrated that we can use artificial language experiments to manipulate artificial signal space proxies to investigate the effects of physical constraints which differ between the speech and sign modalities.

1.1 Aim of this Thesis

In this thesis, I aim to bring together questions and methods relating to both the evolution of language and speech. I specifically aim to ask whether the physical features of a language's modality (speech or sign) affects linguistic transmission and interaction, and ultimately influences the emergence of language and its structure. I will break down the physical differences between the spoken and sign modalities and build hypotheses for how these differences might affect combinatorial structure. I then test these hypotheses using artificial signalling experiments. The overall aim of the thesis is to demonstrate that pressures from the physical features of a signalling modality need to be considered in addition to the effects of cognition, learning, transmission and communication when modelling the emergence of combinatorial structure.

Another aim of this thesis to encourage the interface between computational modelling and psycholinguistic experiments. A lot of the experimental work in this thesis directly tests using real human behaviour conclusions garnered from computational models. The experiments also produce data that can be used in computational models basically out of the box. I worked closely with computer scientists throughout this work and a lot of the data produced from this thesis was used to inform the models in the PhD thesis of Kerem Eryılmaz.

Below, I define the main concepts used throughout the thesis. I will then outline the content of each chapter of the thesis in a road map.

1.2 Terminology

1.2.1 Speech, Language, and Modality

I spend a lot of time throughout the thesis talking about speech, and both the spoken and manual modalities, and about *language* in a broader sense. It is important to define what I mean when I say *language* and what I mean by *speech* and *speech structure*, and how these relate to other modalities.

Language is a human communication system irrespective of its modality. A language's modality is the mode by which a language is produced, which in natural languages can be either speech or sign. The spoken modality uses oral articulators to create auditory signals. The sign (or manual) modality uses gestural means to create visual signals.

This thesis is about the combinatorial structure we see in both language modalities. This can also be called phonology, phonological patterning or sub-lexical structure. Exploring the differences between combinatorial structure in both modalities is paramount to understand *modality effects*, as the best way to understand the effects that a language's modality has is to see what happens in a different modality. Modality is also occasionally used in the thesis to refer to artificial signal spaces used in experiments. It is important to consider modality effects, not only as something which exists in natural language, but also in artificial language experiments as a result of the features of a signalling modality. This is especially true when a modality is being used in experiments which are extrapolating their results to be relevant to natural language.

1.2.2 Iconicity

Besides combinatorial structure, the other main recurring concept in this thesis is *iconicity*. In its broadest sense, iconicity is defined as being a property of a sign when it bears any similarity with its referent. However, the breadth of linguistic behaviour this definition encompasses is not nuanced enough for the needs of this thesis. Iconicity can be broken down into different types of iconicity, which exist on a continuum (Dingemanse et al., 2015). The two main types of iconicity I will refer to throughout are *imagic iconicity* and *diagrammatic iconicity*, though iconicity is not confined to these forms. Imagic iconicity (also sometimes known as absolute iconicity) is when signals resemble their referents, or have some feature in common with their referent. Diagrammatic iconicity (also sometimes known as relative iconicity) is when there is a correlation between the structure of signals and the structure of the meaning space they refer to (see figure 1.1). This definition for diagrammatic iconicity is also sometimes used for compositional structure, which is often defined as being some correlation between how signals are combined and how the meanings they are referring to are combined. However, I think the main difference here is the emphasis on the combination of signals and meanings, rather than a correlation between signals and meanings within a bigger repertoire of individual signals and meanings which is characteristic of diagrammatic iconicity (see figure 1.1). Imagic iconicity can generally be measured using only one signal-meaning pair, whereas diagrammatic iconicity can only be measured with knowledge of multiple signals and meanings and their relationships. I expand more on different types of non-arbitrary structure in Chapter 2.

The major issue regarding iconicity in this thesis is the difference in the iconicity possible using the manual modality in comparison to the iconicity possible with the spoken modality. It is easier to produce iconic signals using the sign modality (Fay et al., 2013). However, imagic iconicity is particularly difficult in the spoken modality. Several recent papers have looked at the differences between how combinatorial structure is affected by the presence of iconicity (Verhoef et al., 2015; Roberts et al., 2015),

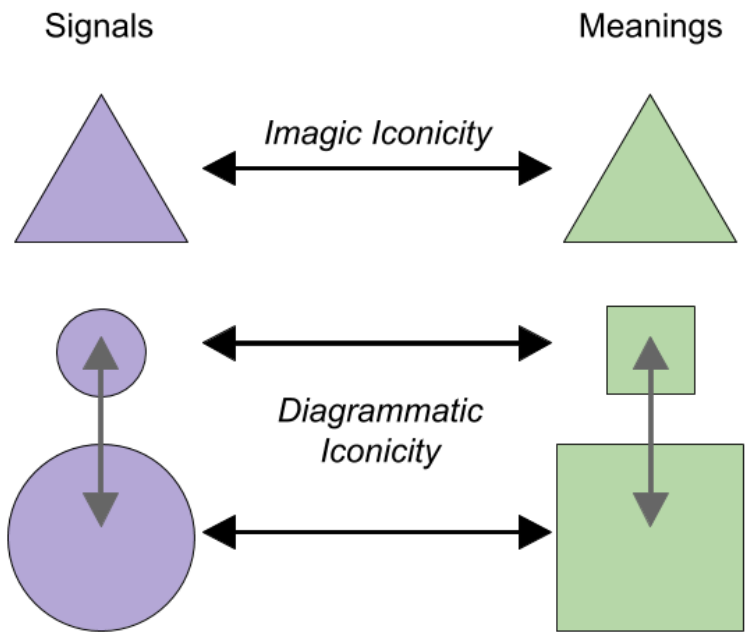

Figure 1.1: Figure adapted from Dingemanse et al. (2015) demonstrating the different types of iconicity. In the imagic iconicity example, the signal and meaning have the same shape which is apparent on viewing only this signal-meaning pair. In the diagrammatic example, a relationship between signals on the left (size) is analogous to the relationship between the meanings on the right.

however, experimental work rarely looks at the effects of different types of iconicity, which is something central to this thesis.

1.2.3 Signals and Meanings

I use the term “signal” throughout the thesis to mean any token used to communicate a meaning, whether it is iconic or otherwise. I use the term “meaning” to refer to any referent used in an experiment for which a signal has been produced. This terminology was used to enable discussion on signal spaces, meaning spaces, and signal-meaning mappings. However, in the literature at large “meaning” is often used to refer to some concept in a human or agent’s brain, rather than the exact instantiation of that meaning, which is usually called a referent. These concepts are difficult to untangle in my experiments, as it is not always possible to see if people have categorised a meaning space in some specific way that would make meanings and referents different during particular tasks. I therefore use the word “meaning” throughout to refer to target meanings (the meaning a signal was specifically made for) and the chosen meanings (the meaning that a participant chose to pair a specific signal with).

1.2.4 Evolution and Emergence

One of the most contentious issues in language evolution is the use of the word *evolution*. Some linguists, most notably Noam Chomsky, claim that evolution can only refer to biological evolution and the field of language evolution should only concern itself with the evolution of the cognitive ability for language (Hauser et al., 2014). This school of thought is fuelled by the assumption that language exists as a biological entity in the brain. Chomsky (2016) states that: “what has evolved is not languages, which do not evolve, [...]. Rather, what has evolved is the capacity for language” (p.200). However, work in the field of language evolution is increasingly looking to cultural evolution as a contributing factor to explain modern linguistic structure. Many scholars argue that linguistic structure comes from small cognitive biases and cultural evolution (Thompson et al., 2016), or sometimes goes as far as to argue that domain-general processes and cultural evolution are sufficient to explain language as it exists today (Christiansen & Chater, 2008). This thesis subscribes to the latter opinion here, that most structure we see in language can be understood as the result of an interaction between domain-general or small cognitive biases, and pressures for learning and communication. However, I do occasionally refer to relevant phenomena which have biologically evolved (e.g. animal communication, the evolution of speech apparatus, etc.). As a result, I sometimes will refer to biological evolution, but I always make it clear when I am talking about biology. For the most part, when using the word *evolution*, I will be referring to the cultural evolution of language.

Within the view of cultural evolution I am subscribing to, it is important to understand that language evolves on 3 timescales that are represented in Figure 1.2. Most work in the field of language evolution looks at the interaction between ontogeny (linguistic development on an individual level) and glossogeny (cultural evolution of the population level). The primary tool for investigating this interaction is cultural learning experiments (covered in the next chapter). Most work investigating this interaction focuses on language learning and use without consideration of physical constraints in language development and use (i.e. modality effects), which is what I focus on in this thesis.

Rather than focusing on the cultural evolution of modern languages (which is the

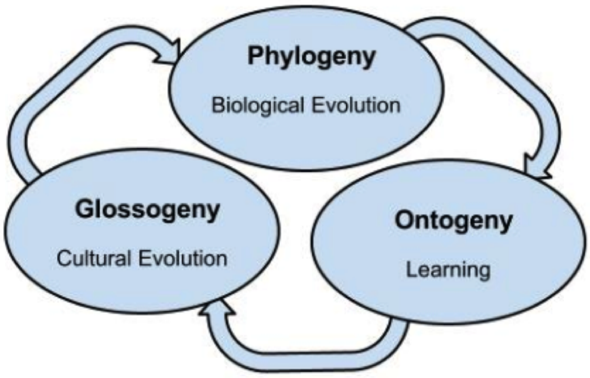

Figure 1.2: The timescales of language evolution as defined by Kirby (2002). Each oval represents its own adaptive system which all interact with one another to create a cycle of ongoing linguistic evolution. It is important that we are able to understand them independently as well as their interaction.

domain of historical linguistics), this thesis concentrates on language origins, how communication systems get bootstrapped, and the initial emergence of linguistic structure. There is work in evolutionary linguistics which has a communication system as a starting point and uses artificial language experiment to study cultural evolution (e.g. Kirby et al., 2008). In this thesis, all of the experiments (except Experiment 1) do not have any system for participants to build on, meaning they need to establish a communication system from scratch (as in a lot of the experimental semiotics literature, Galantucci & Garrod, 2011). As a result, the word *emergence* is generally used, as opposed to *evolution*, in order to refer to the cultural process of structure being bootstrapped and beginning its journey of cultural evolution.

1.3 Thesis Road Map

The thesis is structured as follows:

Chapter 2 asks "where does the structure in language come from?". In the chapter, I will lay out exactly what the structure of language looks like, define the different levels of structure in more detail, and discuss how the different levels of structure might influence each other and co-evolve. I will look at how the definitions I lay out for the levels of linguistic structure relate to both structure in natural language, as well as structure in animal communication systems, and structure which emerges in computational models and experimental work. I will then focus on combinatorial structure, as the main topic of this thesis, and outline existing hypotheses regarding its origins including factors such as learnability, signal space crowding and iconicity.

Chapter 3 discusses modality effects. In this chapter, I will characterise how the spoken and sign modalities differ, and how their differences might affect linguistic structure. The chapter goes on to link modality effects to each of the hypotheses regarding the emergence of combinatorial structure, as outlined in Chapter 2.

Chapter 4 outlines the methods used throughout the thesis and places them in the context of previous work which has used similar methods. I begin by outlining the different experimental structures used throughout the thesis (e.g. iterated learning, individual signal creation, and communication games) and what real-world phenomena they are modelling. I then outline different continuous signal space proxies used in existing experimental work exploring the emergence of combinatorial structure. I go on to outline the main signal space proxy used in the experiments presented in the thesis, detailing its practicalities and its theoretical merits and shortcomings.

Chapter 5 presents Experiment 1, an iterated learning experiment which tests whether the size of a signalling space affects how quickly combinatorial structure emerges. This experiment only alters the size of the signalling space, rather than changing the shape or dimensionality of it in any way, and the signals did not refer to meanings. This experiment shows that in the condition with the smaller signal space, combinatorial structure emerges more reliably than in the condition with the bigger signal space. This chapter deals with the modality effect of size of signal space, with the experimental manipulation modelling the effect of the manual modality having a larger signal-space than than the spoken modality.

Chapter 6 presents Experiments 2 and 3, artificial signal creation experiments which test whether the dimensionality of a continuous signal space in relation to the dimensionality of a continuous meaning space affects the emergence of combinatorial structure. The chapter is specifically interested in the hypothesis that structure will emerge when meaning dimensions outnumber signal space dimensions and that diagrammatic iconicity will emerge when they do not. Experiment 3 learns from problems with Experiment 2, and features a meaning bottleneck and more exaggerated differences between the dimensionality of the signal space and the meaning space. These experiments show strong evidence for the emergence of diagrammatic iconicity when signal and meaning spaces have the same number of dimensions, and evidence for the emergence of structure in signals when the number of meaning space dimensions outnumbers the number of signal space dimensions. This chapter deals with the modality effect of dimensionality of signal space, modelling the pressures imposed by the manual modality having more signal dimensions than the spoken modality.

Chapter 7 presents Experiment 4, an artificial signal creation experiment which tests the effects of different types of iconicity on structure in signals. This experiment uses a continuous signal space which is either used to create signals for meanings which either differ continuously (enabling diagrammatic iconicity) or are discrete (making diagrammatic iconicity more difficult). This experiment again showed strong evidence for diagrammatic iconicity emerging when the meaning space was continuous. Signals had more movement when meanings were discrete, mostly as the result of participants attempting to use strategies with more abstract forms of iconicity. We also found that diagrammatic iconicity made signals harder to discriminate. This chapter focuses on how easy it is to use different types of iconicity in different real world language modalities.

Chapter 8 presents Experiments 5 and 6, an individual artificial signal creation experiment and a communication game respectively. These experiments adopt a meaning space which has less internal structure than earlier experiments, in order to get away from signal structure which mirrors structure in the meaning space, which can be interpreted as diagrammatic iconicity or compositional strcture. These experiments look into the effects of two things: the first is the effect of conventionalisation (the process of signals losing iconicity as the result of repeated use or interaction). The second is the effect of pressures for discrimination, introduced by an expanding meaning space. Signals failed to become less complex in both experiments, which would be predicted by previous work on conventionalisation (Garrod et al., 2007). Signals actually gained complexity later in Experiment 5, which I argue is due to pressures for discriminability. The results from these experiments are compared in order to identify effects of communication. The signals produced with communication were more complex. The results are also compared to results from the existing literature (Garrod et al., 2007) which use graphical representations (drawings) as signals. In my analysis, I treat the graphical signals from Garrod et al. (2007) as analogous to gestural signals, because they are visual, highly iconic and can communicate multiple meaning elements simultaneously. In contrast, the signals from my experiments are more analogous to speech as my paradigm is auditory, restrictive in its production possibilities, and restrictive in its ability to produce signals with high levels of iconicity. In Garrod et al. (2007) signals lose their complexity during communication because of the process of conventionalisation. I account for this not happening in my communication game because the signal space in my experiment (and in the spoken modality) will be saturated by meanings more quickly than when using graphical signals (or the sign modality), which means signals need to increase in complexity in order to maintain their distinctiveness.

Beyond complexity in signals, another way to measure the process of conventionalisation is to measure the amount of iconicity in signals. Therefore, chapter 8 also presents Experiment 7, an online playback experiment where naïve participants on the internet were asked to listen to signals from both Experiments 5 and 6 and select which meaning they thought the signals referred to. If naïve participants can identify the correct referent for a signal, then that signal can said to have imagic iconicity. Iconicity increased over time during Experiment 6, suggesting signals were not being conventionalised. Iconicity stayed the same throughout Experiment 5. These results are again compared to Garrod et al. (2007) in order to understand what they may tell us about modality effects.

Chapter 9 presents Experiment 8, an artificial signal creation experiment which tests the effects of signal duration on structure in signals. This experiment is a replication of Experiment 5 but signals were only allowed to be 1 second long. The effects of duration constraints on signals haven't been systematically investigated in the artificial language literature, though some experiments in the literature impose duration limits while others do not. Physical and functional pressures might affect how long signals in the real world can be. For example, speech is constrained by breath, which the sign modality is not. In this experiment, signals produced with the time constraint were much simpler than those produced in Experiment 5.

Chapter 10 discusses and concludes the contents of the thesis, as well as suggesting ideas for future work that have grown from the work presented throughout.

CHAPTER 2

Where does structure come from?[1]

Where did the structure in language come from? This is the main question in the field of language evolution. In this chapter, I will define the different levels of structure in language and how they can be characterised in a way that provides useful comparisons between languages in the real world, structure in animal communication, and structure in artificial languages produced in experimental studies. I will then go on to discuss important evidence that illustrates how structure on different levels can influence each other. Further, I will briefly discuss which level of structure was likely to have come first, which of course has implications for how structure at different levels emerged. As the primary focus of the thesis is combinatorial structure, the chapter will close by explaining current hypotheses on the origins of combinatorial structure specifically.

2.1 Duality of Patterning

Duality of patterning, the property of language that it is made up from combinatorial and compositional structure (defined below), is one of the original design features proposed by Hockett (1960). Hockett's (1960) original design features are a list of universal characteristics, which are said to be present in every language in the world, and also unique to humans. However, whether duality of patterning should be defined as being "universal" is controversial. Most notably, data from rural sign languages, that have only recently emerged, show that these languages can be fully expressive with-

[1]Some material from this chapter is published in Interaction Studies (Little et al., 2017d).

out having a level of combinatorial structure (e.g. Al-Sayyid Bedouin Sign Language (ABSL), Sandler et al., 2011). Also, whether duality of patterning is uniquely human is also debatable as there is some evidence of duality of patterning in monkey calls. I will talk about both rural sign languages more thoroughly in the next chapter, and animal communication more extensively in the sections below, but it is first important to identify exactly what I mean by combinatorial and compositional structure.

2.1.1 Combinatorial Structure

Combinatorial structure is the combination of meaningless building blocks (e.g. phonemes) to make meaningful units (e.g. words and morphemes) (de Boer et al., 2012). More precisely, combinatorial structure is when two or more signals combine to make a composite signal, the meaning of which does not correspond to the sum of its parts (Scott-Phillips & Blythe, 2013).

Assessing whether something is combinatorial structure is not always clean-cut in language. A phoneme that has been proved meaningless in other contexts can sometimes stand alone to denote meaning, as well as contribute to a meaningful composite signal in the form of combinatorial structure. For example, the phoneme /ə/ in English can have meaning when used as the indefinite article in constructions such as /ə ˈbʊk/[2] (a book), but can also be a meaningless building block in words such as /nəːs/ (nurse). In real world languages, phonemes are defined as being units that, when exchanged for one another, can change the meaning of a word. For example, /hoːm/ (home) and /hoːn/ (hone) differ in their final phonemes (/m/ or /n/), which changes the meaning of the word. These words are called *minimal pairs*. If a language has minimal pairs, then it can be said to have combinatorial structure. However, some may argue that a language does not have to have minimal pairs to have combinatorial structure. Meaningless units can exist and be reused within a larger linguistic system without ever appearing in a contrastive context. These contrastive contexts (or minimal pairs) are a lot more likely with fewer building as a starting point, and therefore systems with many building blocks can possibly have combinatorial structure without minimal pairs. Systems with and without minimal pairs can also be subject to rules or statistical tendencies regarding the order in which phonemes can be combined. These rule are called *phonotactics*.

[2]The IPA transcriptions throughout this thesis are representative of the author's dialect (Teesside English).

2.1.1.1 Combinatorial Structure and Indissociable Features

Phonemes in natural languages are made up from feature bundles. In the home/hone example, the /m/ and /n/ are the phonemes that change the meaning of the word. However, there is only one feature different between the /m/ and /n/, which is their place of articulation. They are both voiced nasals, but the /m/ is bilabial and the /n/ is alveolar. These features that make up the phonemes in spoken language then, are themselves combined to make phonemes. However, these features are *indissociable*, which means that they can't exist alone or outside of a combination with other features. A phoneme must simultaneously have voicing, manner of articulation and place of articulation. As not combining these elements is physiologically impossible, we cannot call this feature of language special or symptomatic of combination on a cognitive level. It is therefore also difficult to argue for meaning being attached to indissociable features in spoken languages. However, some features in the artificial signal spaces used throughout this thesis have indissociable features that participants do attach to meaning, which is discussed in Chapter 6. Further discussion of how combinatorial units being indissociable relates to signed languages is discussed in Chapter 3.

2.1.1.2 Bare Phonology

Meaningless building building blocks can also combine to make units that also do not have an explicit meaning. This has been referred to as *bare phonology* by Fitch (2010). Examples of bare phonology include music or birdsong, but there also exist examples in the experimental literature (Verhoef et al., 2014).

2.1.2 Compositional Structure

Compositional Structure is the combination of meaningful units (such as morphemes or words) to make meaningful units (phrases and sentences). More precisely, the definition of compositional structure also involves that the way or order in which meaningful units are combined can contribute to the meaning. The meaning of a sentence or a compound-word is often guided by not only the meanings of the words or morphemes involved, but also the order in which they occur. For example, a "fish pond" (a pond with fish in it) has a different meaning to "pond fish" (a type of fish that dwells in ponds). However, the order of combined elements isn't always important to their composite meaning, for example, "the Belgian-Dutch border" and "the Dutch-Belgian border" have the same referent.

2.1.2.1 Compositional Structure as Iconicity

Compositional structure can be defined by structure in a signal correlating with the structure of a meaning space. For example, the noun phrases "red book" and "red ball" both have the word "red" in them, which correlates to redness seen in these items in a meaning space. However, "red book" and "blue book" do not have redness in common, but they do correlate regarding books. As a result of this correlation between signal and meaning spaces, there is sometimes an overlap between definitions of compositional structure and diagrammatic iconicity, which is particularly salient to some of the examples throughout this thesis, particularly in Chapter 6. However, as the focus of this thesis is combinatorial structure, I will not indulge in an overly extensive discussion of compositional structure as iconicity here, beyond raising this issue now, in order that I can flag when this overlap becomes an issue later in the experimental parts of the thesis. For an extensive discussion on compositional structure as iconicity see Haiman (1985).

2.2 Structure in different contexts

The definitions above are designed to be broad, so that we can use them to describe behaviour across different contexts. However, all the examples above relate to structure in spoken natural human language. This thesis is about modality effects, so it is important to clarify how structure is characterised in signed languages. The thesis also uses artificial language experiments, which act as a proxy for real world languages, so it is important to understand how combinatorial and compositional structure manifest in experimental contexts. Furthermore, it is useful to consider whether and/or how structure exists in animal communication systems. If analogous structures do exist in the animal world, we can look at what affects their emergence and how structure can be characterised in the animal world, to see if we can learn anything about analogous structure in human languages.

2.2.1 Structure in Signed Languages

The vast majority of signed languages have duality of patterning, which is broadly comparable to the same structures in spoken languages. Signed languages have compositional structure in the same way that spoken languages do; words and morphemes can be combined to make bigger meaningful structures, the meaning of which can also be affected by word order and other grammatical functions. Signed languages also have the ability to present morphemes simultaneously, which is not possible in spoken languages (Meir et al., 2010).

Signed languages also have combinatorial structure. Stokoe (1960) was the first to identify the individually meaningless features that can be combined into meaningful units. In the same way that voicing, place of articulation and manner of articulation are features that combine to make phonemes in spoken language, features combine to make phonemes in signed languages, including hand shape, location of hand, movement of a hand, and sometimes facial expression. Combinatorial structure is measured in the same way that it is in spoken language, via the use of minimal pairs. However true minimal pairs are much rarer in signed languages (Sandler, 1996), possibly the result of signed languages having bigger phonemic inventories. As mentioned above, there are also signed languages that have been discovered to not have any minimal pairs (e.g. Al-Sayyid Bedouin Sign Language, Sandler et al., 2011) and there are also several differences between combinatorial structure in signed languages and spoken languages, but these are discussed more extensively in Chapter 3.

2.2.2 Structure in Artificial Language Experiments

As I said in the introduction, investigating where structure came from is the main focus of the field of language evolution. One of the main avenues of research in evolutionary linguistics, including the work done in this thesis, is the use of artificial language experiments. As a result, we need to understand how different levels of structure manifest themselves in artificial languages.

Experiments in the language evolution literature almost exclusively focus on the emergence of structure on one level, either compositional or combinatorial. Studies focusing on the emergence of compositional structure typically start with a set of discrete building blocks (typically letters) that are arranged into syllables that have some kind of simple phonotactic rules inspired by real world languages, e.g. consonant–vowel (CV) pairs or CVC syllables. Examples of such studies include, but are not limited to, Kirby et al. (2008), Kirby et al. (2015) and Silvey et al. (2015). These studies find that syllables in the artificial words come to denote elements of the meanings they describe. Such compositional structure is easy to measure by correlating elements of the signal space with elements of the meaning space. The artificial languages that emerge in these experiments are only very small (less than 30 words). As a result, it is very difficult to know if the order of elements affects meaning. Further, it is impossible to observe the emergence of structure on a combinatorial level in these experiments because the systems start with discrete building blocks and phonotactic rules. Also, as these studies use letters, then participant's pre-existing knowledge of how letters (or phonemes) combine will affect their behaviour.

In order to investigate the emergence of combinatorial structure using artificial language experiments, methods have been developed that do not start with a set of pre-existing building blocks, but a continuous signal space. These methods include the use of slide whistles (Verhoef et al., 2011) and graphical interfaces (Galantucci, 2005), and will be reviewed more extensively in Chapter 4. Measuring combinatorial structure, in the terms defined in the current chapter, in artificial languages produced using these methods is difficult for several reasons. First, when meanings exist in an experiment, the structure of signals tends to mirror that of the structure of meanings, either through compositional structure or iconicity, making it difficult to identify units as being meaningless. One way that some studies have got around this problem is to minimise the amount of internal structure in the meaning space or to disrupt the ability to use iconicity (e.g. in Verhoef et al., 2015, or Roberts et al., 2015). Secondly, only very limited numbers of signals can be produced using these experiments in a short amount of time, making the emergence of a neat system of minimal pairs unlikely. Instead, studies attempt to measure the reuse of units and groups of units in signals. Some of these methods are reviewed in chapter 4. Such measures can help us answer whether signals or repertoires are structured, or if discrete units exist within the signals, but not necessarily whether these units are meaningless. As a result, to know whether structure in artificial signals is combinatorial structure or otherwise, we can use subjective means either by qualitative analysis by the experimenter, or by asking the participants themselves using post-experimental questionnaires, something that has been adopted in the current thesis.

2.2.3 Structure in Animal Communication Systems

Does structure exist in animal communication systems? And is the structure we see combinatorial or compositional? These questions may help us understand what causes similar structures in human languages. At least one recent review (Collier et al., 2014) has drawn on evidence from animal calls to determine whether combinatorial structure could have preceded compositional structure, or vice versa. Peter Marler has previously outlined different types of call combinations in animals calls, making a distinction between "phonological syntax" and "lexical syntax", which are analogous to combinatorial and compositional structure respectively (Marler, 1998). Marler argues that animal calls have yet to be found to display anything that we might be able to call compositional structure. Since then, Scott-Phillips & Blythe (2013) have argued compositional structure is rare in the animal kingdom, the central argument being that only humans are capable of compositional structure, as only we have the inferential abilities to unpack novel signal-meaning mappings. As animal communication systems are so plentiful and varied, I will briefly discuss 3 examples here which I believe are relevant to this thesis. I will write about birds, honeybees and monkeys.

2.2.3.1 Birds

Birdsong is generally considered to be "bare phonology" as there is no one-to-one mapping between units in the song and semantic content. However, Suzuki et al. (2016) have shown that Japanese great tits (*Parus minor*) possibly have compositional structure in their vocalisations. They found that a combination of notes that generally denote 'scan for danger' and 'approach the caller' meant that the birds did both. Further, when the notes were reversed so that the note meaning 'approach the caller' occurred first, the birds did not do both, which indicates that it was not only the combination, but the way in which they were combined that results in the behaviour. The sequence of notes meaning 'scan for danger' is made up from three smaller notes that possibly indicates combinatorial structure, as these notes are also reused in other note-combinations used when the birds are mobbing predators.

2.2.3.2 Honeybees

Honeybees (*Apis mellifera*) communicate via a system of movements called a waggle dance. The dance differs on 2 dimensions; the angle of the dance, and the duration of the dance (Von Frisch, 1967). These dimensions represent where flowers are in respect to two meaning dimensions; the angle with the sun and the distance respectively. Some varieties of bees also indicate the quality of the flowers using the intensity of the dance (Waddington, 1982). The signals have a diagrammatic iconic relationship with the meanings they are communicating, which might even be described as compositional, in that the meaningful elements in the signals are combined. However, the signal elements in waggle dance are indissociable in the same way that phonemic features are. A dance must have an angle, and it must have a duration. As a result, the signal elements are not flexible in how they can be combined. In the current thesis, there are signal spaces with indissociable elements. Some experiments have signals that have pitch, volume and duration simultaneously, and meaningful correspondences to each. Looking at the bee example, it is clear how such a system might be arrived at by humans given the same apparatus (a dance) and communication task (finding flowers a certain direction and distance away). It is possible then, that the use of indissociable elements in a communication system may depend massively on the signal space or modality being used, as well as how this maps to a given meaning space. These themes are raised throughout the thesis.

2.2.3.3 Putty-nosed and Campbell's monkeys

The most oft-cited example of an animal with combinatorial structure is the Putty-Nosed monkey (*Cercopithecus nictitans*). These monkeys have a communication system made up from two different vocalisations: pyows and hacks. These calls, as summarised by Arnold & Zuberbühler (2006), are used to communicate different environmental states. If there are eagles present (a predator of the monkey), they make a 'pyow' call, if there are leopards present (also a predator), they makes a 'hack' call. If, however, there is an absence of food, they use a combination of the calls: 'pyow–hack'. An absence of food is not a combination of the presence of leopards and eagles, so this combination is not compositional, but combinatorial (de Boer et al., 2012). Schlenker et al. (2016) have since made the argument that 'pyow-hack' may indeed be compositional if one takes pyow to mean 'general alarm' and hack to mean 'non-ground movement', and then 'pyow-hack' to mean the movement of the monkey group (whether as the result of a lack of food or otherwise). However, both Scott-Phillips & Blythe (2013) and Collier et al. (2014) argue instead that this call is, in fact, holistic and not combinatorial as experiments have demonstrated that the monkeys interpret this sign idiomatically (Arnold & Zuberbühler, 2012). It is worth noting here that the definition of combinatorial used by Scott-Phillips & Blythe (2013) is more similar to the one I use for "compositional" throughout this thesis. However, regardless of the semantic interpretation of the calls, it is true that these monkeys combine two signals to create a third meaningful signal.

There is also some evidence for compositional structure in another type of monkey, the Campbell monkey (*Cercopithecus campbelli*). The Campbell monkey has three signal elements: 'krak' denoting a leopard, 'hok' denoting an eagle and '-oo', which

Collier et al. (2014) argue acts like a suffix that denotes a disturbance. Indeed, 'krak-oo' denotes a general disturbance and 'hok-oo' denotes a disturbance in the canopy. Whether a conscious association between signal elements and meaning elements is made by the monkeys though, is the important factor in determining whether the calls are compositional or combinatorial structure, which is obviously difficult to establish.

The combination of calls in both species most likely came about as they only have the ability (either physically or cognitively) to produce two different signals, so when they need a necessity for a third call arose, they only have the option to combine the 2 calls they already had. This is a very simplified version of the pressure of signal space crowding affecting the emergence of combinatorial structure that I will discuss in the hypotheses section below.

2.3 Combinatorial and Compositional Structure: Which Came First?

Combinatorial and compositional structure are not autonomous. Structure on both levels affects structure on the other. For example, Wedel et al. (2013) demonstrated that low functional load on the compositional level can lead to a loss of phonemic contrasts on the combinatorial level. Further, phonological cues (by definition on the combinatorial level) can help the acquisition and categorisation of word classes (by definition on the compositional level) (Dingemanse et al., 2015). As structure on both levels are co-evolving with each other in modern language then, it is sensible to assume that they co-evolved and co-emerged together when our ancestors first started to use language. However, there is an ongoing controversy in the literature surrounding the issue of whether combinatorial or compositional structure came first.

There are examples of human languages that have existed without a level of combinatorial structure (such as emerging sign languages, which are covered in the next chapter). Collier et al. (2014) use their interpretation of Campbell's monkey calls having compositional structure to make the case that syntax came before phonology. However, the case of Putty-nosed monkeys demonstrates that the opposite is just as possible. Tria et al. (2012) use a computational agent-based model where successful compositional lexicons emerge faster than combinatorial ones, suggesting that compositional structure may have emerged first under the specifically functional needs of communication. Others, such as Fitch (2006) and Verhoef et al. (2014) have argued for the possibility that combinatorial structure could have emerged first as some kind of musical protolanguage without meaningful units (see section on bare phonology above). However, isolating examples where compositional structure exists without combinatorial structure or vice versa is only evidence to suggest that syntactic structure can precede phonology, or vice versa, rather than suggesting that it must. There are all sorts of factors that might make one scenario more likely than another. In this thesis, I argue that whether combinatorial or compositional structure came first is tied up with the modality being used and features of the physical signal space. A lot of the hypotheses for how and why combinatorial structure originated and emerged are tied up with features of language modalities. I list the hypotheses below, and outline how they interface with modalities in the real world in the next chapter.

2.4 Hypotheses about Combinatorial Structure

This thesis is about the emergence of combinatorial structure. There is already a substantial body of work on this question, so in this section I will break down the existing hypotheses about what pressures contributed to the emergence of combinatorial structure. I will first cover learnability as a driver for combinatorial emergence, then signal space crowding and finally discuss iconicity as a possible factor. It is important to note that these hypotheses are not completely separable from one another, nor are they in any way mutually exclusive. I support a big picture account in which all of these mechanisms have a role.

2.4.1 The Pressure of Learnability

Language adapts to the be learnable. This has been shown extensively in both computational work (see Brighton et al., 2005, for a review, and Oudeyer, 2005, for a combinatorial-specific example), and in experimental work (see Kirby et al., 2008, and Verhoef et al., 2014, for a combinatorial-specific example). Essentially, the reuse of discrete units from established signals enables efficient encoding of information, making new signals easier to learn. If a signal is easier to learn, it is more likely to survive in a population.

Pressures for learnability can be characterised by a learning bottleneck. Learning bottlenecks result from individuals only being able to learn a set number of signal-meaning pairs, either as the result of having a limit of their memory capabilities, or because of a poverty of stimulus, where they only get exposed to a subset of possible signals.

Verhoef et al. (2014) tested the learnability hypothesis experimentally. Participants used slide whistles to reproduce signals that were part of an "alien language". They learnt a set of 12 whistles and were then asked to reproduce all 12. The bottleneck in this process was simply tied to the difficulty of the task. It is reasonably hard to remember 12 whistle signals, and so if participants could not remember all 12 they were instructed to make up a new signal that they felt "fit with the language". Their reproductions were then passed to a new participants to learn. This process was repeated over 8 generations. Verhoef et al. (2014) showed that structure emerged in these signal repertoires, in the form of reused building blocks, simply as a result of this process of learning and transmission. Del Giudice (2012) found similar results with an iterated learning experiment but using continuous signals that were graphical rather than auditory.

The majority of the focus in this thesis is not on learnability, but on the hypotheses listed below. Learnability has been well demonstrated as an effect in previous work, and the current thesis assumes learnability as one of the main drivers for the emergence of combinatorial structure. Learnability is also not completely separable from the other 2 sets of hypotheses below. Memory constraints are very tied to crowding in signal spaces if one thinks about the signal space in terms of the number of signals an individual can remember. Further, iconicity is not separable from learnability as iconic signals may affect the age or ease of acquisition (Monaghan et al., 2014).

2.4.2 Signal-space crowding

Hockett's (1960) hypothesis for what caused the emergence of combinatorial structure is that it is the result of pressures imposed when the number of vocabulary items increases, as a language would run out of ways to create meaningful distinctions. This overcrowding then causes a requirement for a more efficient way to create new word forms to maintain distinctions while retaining expressivity. He proposes combinatorial structure as the more efficient way. This hypothesis relies on the assumption that meaningful holistic signal units are reanalysed to be the meaningless units that make up combinatorial structure. This hypothesis was originally formulated by Hockett (1960) in terms of the number of meanings in a system, however, the size of a signalling space will directly affect how quickly a language runs out of ways to create meaningful distinctions because of how quickly that signal space gets crowded. A good formulation, then, would rather be the ratio of meanings to possible distinctions in the signal space.

Hockett (1960) was not the only one to to come up with such an idea. However, some have observed that such a mechanism is not unique to language. Abler (1989) proposed the particulate principle. This principle argues that "the only route to unbounded diversity of form and function is through a combinatorial hierarchy in which discrete elements, drawn from a finite set, are repeatedly permuted and combined to yield larger units" (Studdert-Kennedy & Goldstein, 2003, page 235). This principle holds for many physical systems in the real world; in chemistry, genetics and music, among many others.

The most obvious thing that will affect how quickly a signal space becomes crowded is the size of a signal space. The bigger a signal space, the more signals that can exist in it without it getting crowded. This straightforward prediction is already investigated in some experimental work. Roberts & Galantucci (2012) conducted an experimental semiotics study in which participants communicated about different animal silhouettes. They found no correlation between the number of animals communicated at the end of the experiment (how crowded the signal/semantic space was) and the amount of combinatorial structure in signals. However, there were other confounds in this analysis as the number of meanings participants saw was dictated by how successful they were in the experiment.

Further, since structure emerged rapidly in Verhoef et al. (2014), where there were only 12 signals, then this may be evidence to suggest that crowding in the signal space may not have such a strong effect as pressures for learnability. However, as I mention above, memory is one constraint that may make the signal space smaller at a cognitive level. That is, only being able to remember a set number of signal units, and therefore having to recombine them, is essentially the same process as only being able to distinguish a set number of signal units, and therefore having to recombine them. In chapter 5, I replicate Verhoef et al. (2014) with a condition where the signal space is smaller to see if the size of the signal space may also have an effect.

Further to the above, other things can affect the mechanism of how quickly a signal space gets crowded. Here, I have isolated 3 of these features: dimensionality, noise and nonlinearities in the signal space. Further, signal duration and rapidity of fading are identified as factors that may have their own effects on crowding within signals, signal spaces and also generate constraints of memory. I have given each of these features its own section below.

2.4.2.1 Signal Space Dimensionality

The number of dimensions that a signal space has, is the amount of ways that meaningful distinctions can be made using that signal space. Dimensionality is strongly linked to the size of the meaning space, as the more dimensions a space has, the more distinctions can be made within that space. This dimensionality might amount to what we intuitively think of as dimensionality in mathematical terms, for example a toy example that is sometimes used is the two-dimensional vowel space, which can be plotted with regards to the physical position of the tongue when producing vowels, or by plotting vowel formants, with the first vowel formant on one axis, and the second vowel formant on the second axis. Of course, the vowel space has more than 2 dimensions when one starts probing further (roundedness, voice quality, fundmental frequency, etc.), but its dimensionality is still relatively more intuitively imagined than that of other signalling modalities. The ways in which signals can differ do not always map to spaces where we can intuitively think about dimensionality, for example consonants can differ by voicing, place of articulation and manner of articulation. These phonemic features would all count as dimensions within the current definition.

Meaning spaces also have dimensions. Meaning dimensions are the number of ways that meanings can differ. Accordingly, it is possible to map signal dimensions to meaning dimensions. For example, having signals that differ on the dimension of volume map to meanings that differ on the dimension of size with louder signals referring to bigger meanings. This is an example of diagrammatic iconicity, as defined in the introduction. When such a mapping happens, it creates crowding on the signal dimension when there is a lot of gradations of difference along a meaning dimension. Previously, Gasser et al. (2010) hypothesised that more crowded semantic spaces will benefit from words that are more distinguishable from one another. That is, strategies of diagrammatic iconicity will not be beneficial in such situations, meaning that strategies that provide more distinguishable signals (such as combinatorial strategies) may be a good resolution. This again amounts to the same hypothesis about crowded signal spaces being unsuited when there is a pressure for expressivity.

Above, I have discussed pressures created by the dimensionality of a signal space if there a one-to-one mapping with the dimensionality of a meaning space. But what happens when there is no one-to-one mapping available between dimensions? de Boer & Verhoef (2012) empirically investigate the emergence of combinatorial structure as a result of the number of meaning dimensions outnumbering the number of signal dimensions. Using a mathematical model, they found that when there is the same number of signal and meaning dimensions, then signals emerge that have diagrammatic iconicity. However, when the meaning dimensions outnumber the signal dimensions, then structural strategies emerge. Whether this structure is combinatorial or compositional relies on the specific mappings involved. I experimentally test this model in Chapter 6.

There are a couple of references in this dimensionality section to issues of diagrammatic iconicity. How iconicity relates to the emergence of combinatorial structure is discussed in the iconicity section below.

2.4.2.2 Noise in the Communication Channel

Noise is any kind of interference in the communication channel that prevents the faithful transmission of a signal. Shannon (1948) was the first to mathematically investigate noise having effects on the transmission of signals. It is a well established hypothesis that pressures involved with producing signals in a noisy environment facilitated the emergence of combinatorial structure, precisely because it makes the signal space more crowded, not with signals, but with noise. Though it is worth noting that adding noise to a signal space is mathematically identical to keeping the noise constant and making the signal space smaller. Hockett (1960) himself wrote: "There is a practical limit, for any species or any machine, to the number of distinct stimuli that can be discriminated, especially when the discriminations typically have to be made in noisy conditions." (p.12). If a signal is to be robust against noisy conditions, then there is a limited number of signals that can be produced before they need to start being combined. Nowak et al. (1999) explored this idea empirically using a mathematical model, showing that a noisy communication channel can have an effect of the emergence of combinatorial structure. They demonstrated that with a noisy channel, it was not useful to add more unique signals from a signal space because of what they call an "error limit". That is, having more signals in the signal space decreases communicative success because discrimination is impeded by noise. As a result, the signals need to be combined in order to generate more signals that are resistant to noise.

The argument that sequential strategies make signals more robust in noisy situations has since been explored by several computational models. Zuidema & De Boer (2009) modelled trajectories with temporal structure and measured the fitness of these trajectories by how perceptually distinctive they were, or how robust they were to noise. They found that, indeed, pressures for robustness to noise does drive a system to combinatorial structure. Tria et al. (2012) used an agent based model to investigate the emergence of combinatorial structure, but instead of using trajectories in a continuous signal space as Zuidema & De Boer (2009) do, they started with strings of discrete symbols. The emergence of combinatorial structure again depended on the levels of noise in the communication channel.

2.4.2.3 Nonlinearities in a signal space

Another feature of a signalling space, which is related to dimensionality, but not the same, is non-linearity. What I mean by non-linearity is where there is not a constant mapping between the distances between signals in production and the distances in signals in perception. This is particularly salient in speech where there is often more than one way to produce sounds which are perceptually identical. Janssen et al. (2016) studied the effect of ease of articulation on a set of continuous slide whistle signals, produced digitally using a track pad or mouse. They conducted communication games in iterated chains and found that when non-linearity was present in the signal-space, participants took advantage of stable regions. Janssen et al. (2016) argue that these non-linearities might be at the root of combinatorial structure. It is possible that these nonlinearities could be interpreted as noise in the signal space, as they create perceptual difficulties in some parts of the space, which perhaps would lead them to contribute to the emergence of combinatorial structure.

2.4.2.4 Signal Duration and Rapidity of Fading

The duration of a signal and how quickly a signal fades are connected, but are not the same thing. The duration, as defined in this thesis, is simply the amount of time it takes to produce a signal. A signal with a very small amount of information being drawn out across time could have the same production duration as a signal with a lot of information in it being produced with speed, or indeed, a signal with a small amount of information but a lot of redundancy. Rapidity of fading, on the other hand, is how long a signal lasts once it has been produced. Auditory signals, for instance, are likely to disappear as quickly as they are produced, which was a design feature of speech identified by Hockett (1960). A written signal has the potential to last forever. There is an overlap between these two phenomena, because the rapidity at which a signal fades can sometimes be extended by extending a signal's production time. For example, an alarm can be sounded for a short or extended period; if the alarm goes on for longer, it is both extended in its production, and extended in its fading. However, despite this overlap, the potential effects on combinatorial structure from these phenomena are slightly different. Accordingly, I will address them under separate headings here.

2.4.2.5 Signal Duration

Restricting the duration of a signal may speed up how quickly combinatorial structure emerges. A pressure for shorter signals means a pressure for information density, efficiency in communication and a pressure against the amount of redundancy in signals. Combinatorial structure (and compositional structure) is a great method to get an efficient system to have information conveyed succinctly and efficiently with a finite number of building blocks. Further, redundancy has the potential to detract attention from meaningful elements in a signal, which could interfere with the process of analysing meaningful chunks from holistic signals.

On the other hand, constraints on duration of signals could increase the (evolutionary) time it takes for combinatorial structure to emerge. When signal durations can be longer, the time dimension can be used to counter crowding in a signal space (discussed above). The time dimension is then the thing that facilitates the emergence of combinatorial structure rather than restricting it. Accordingly, it may be that longer signal durations may facilitate combinatorial structure to emerge, if there are other things crowding the signal space. I am not aware of any empirical work on this topic, and therefore I have conducted a small experiment summarised in Chapter 9.

2.4.2.6 Rapidity of Fading

Rapidity of fading could affect the emergence of combinatorial structure because of memory constraints. The slower a signal fades, the more time there is to take the signal in and remember it. If a signal fades quickly, people have less time to process the signal. As we have seen above, combinatorial signals are easier to learn and are more faithfully transmitted. As such, having less time to process a signal may be a driving force to make it more combinatorial. Galantucci et al. (2010) conducted an experiment where in one condition, signals disappeared as soon as they had been produced. In the other condition, signals faded slowly, lasting for 2.5 s before disappearing. In the fast fading condition, participants adopted more combinatorial structure.

2.4.3 Iconicity

An increasing body of literature is emerging on the hypothesis that iconicity may inhibit the emergence of combinatorial structure (Goldin-Meadow & McNeill, 1999; Roberts et al., 2015; Verhoef et al., 2015). When a signal has an iconic relationship with its meaning, this creates a signal with high referential efficiency. Roberts et al. (2015) argue that when a language is first emerging, this referential efficiency may be more beneficial than signals with high transmission efficiency, they term this the Reference-Before-Transmission Hypothesis. As I have argued above, transmission efficiency can come from combinatorial structure. Goldin-Meadow & McNeill (1999) make the very similar argument that communication systems that can rely on the strengths of iconicity do not need to rely on the strengths that combinatorial structure brings. These arguments rely to some extent on the idea that there may be a mutually exclusive relationship between combinatorial structure and iconicity. This is of course not the case; much combinatorial structure can have iconic properties, such as sound symbolism. Further, iconicity that enabled a sign to become bootstrapped within a communication system may become dormant after a process of conventionalisation. However, this does not mean that the iconicity has gone (Roberts & Galantucci, 2012). After a sign has conventionalised, people may be more likely to reanalyse elements of a signal that they perceive to be meaningless because this iconic mapping is no longer clear to them.

Several studies have empirically investigated the relationship between being able to iconically represent meanings and the emergence of combinatorial structure using artificial language paradigms with continuous signal spaces. Roberts & Galantucci (2012), in the same study mentioned above, where participants communicated about different animal silhouettes, found an inverse relationship between iconicity present in signals and the amount of combinatorial structure. Since then, Roberts et al. (2015) conducted an experiment to further test this idea that iconicity can inhibit the emergence of combinatorial structure. In this study, also a communication game, meanings were either wavy lines, which could be represented through imagic iconicity as the signalling apparatus could produce wavy lines, or circles that were various shades of green, which were much more difficult to be iconically represented. The experiment showed that indeed, the signals used for circles were made up from repeated, combinatorial elements, while the lines retained iconicity without the introduction of structure within the signals.

Verhoef et al. (2015a) tested whether iconicity affected the emergence of combinatorial structure using the slide whistle signals in an iterated learning experiment. One condition was the same as Verhoef et al. (2014), mentioned above, only the signals were attached to meanings. These meanings were designed to be novel objects for which participants had no existing labels. The meaning space had no internal structure, but it was possible to create iconic mappings between the meanings and the signals. In one condition, whistles and meanings kept their pairings between generations. So the signal that one participant generated for a meaning was presented with the same meaning to the next participant in the transmission chain. In the other condition, one participant's signals were passed to the the next participant in the chain, but they were paired to different meanings randomly. In the first condition, iconicity was much easier because signal-meaning mappings were the same for producers and learners. In the second condition, iconicity was difficult because signal-meaning mappings were interrupted with every turnover of a generation. They found that combinatorial structure emerged more slowly in the condition with consistent signal-meaning pairs, suggesting that iconicity was inhibiting the emergence of combinatorial structure.

2.4.3.1 Types of Iconicity

Further to the studies above, which tend to treat iconicity as one broad concept that has blanket effects, I think it is important to discuss the possibility that different types of iconicity might affect combinatorial structure in different ways. These have already been discussed above in the section on signal space dimensionality. Diagrammatic iconicity can encourage crowding of a signal space, which has knock on effects to combinatorial structure. Further, more abstract forms of iconicity, which rely on metaphor or individual features of meanings, may conventionalise in different ways, which will have a knock on effect on the emergence of combinatorial structure. These issues are discussed further in chapters 5 and 6.

2.5 Conclusion

In this chapter I have outlined definitions for the different levels of structure in language and described how these definitions relate to communication systems in different contexts, such as in artificial language experiments and animal communication systems. The most important part of the chapter, however, is the outlining of different hypotheses about the emergence of combinatorial structure. These hypotheses create the backbone of the thesis, and will be what I aim to test using experimental methods throughout. In the next chapter, I discus how these hypotheses relate to real world linguistic systems that use different modalities. This, I hope, will make clear the implications of these ideas to languages in the real world.

CHAPTER 3

Modality effects

Modalities in real world languages (speech and sign) have several differences in how they produce and convey linguistic signals. This is also true of the artificial signalling modalities used in experimental work. These differences can have effects on the structure we see in the systems produced in these modalities. Here, I will explore how language modalities differ and how their differences might affect linguistic structure, specifically on the level of combinatorial structure. I will also occasionally refer to artificial signalling modalities used in experiments and how artificial modalities relate to the modality effects I identify. Further, I will go through each hypothesis outlined in chapter 2 and relate these back to the issues discussed in this chapter.

3.1 Modality Effects

Despite the assertions of many early linguists (Bloomfield, 1933; Hockett, 1960; Sapir, 1921), signed languages are as complex, expressive and productive as spoken languages. Further, humans are just as adept at acquiring signed language as they are spoken language with the same rate of acquisition (Newport & Meier, 1985). There is now a large literature exploring similarities between signed and spoken languages, and it is uncontroversial to assert, especially at the level of cognition, that there are far more similarities between linguistic systems using either modality than there are differences. However, given that the means of production are so different between signed and spoken languages, there are still many ways in which the systems differ.

These differences have the potential to have knock-on effects on language evolution and linguistic structure.

Previously, Meier et al. (2002) edited a monograph exploring how the manual modality might affect the structure we see in signed languages. This chapter will briefly cover some of the modality effects outlined in Meier et al. (2002), with some additional ones that are explored in this thesis. Where this chapter departs from the impressive breadth of Meier et al. (2002), is that modality effects here are specifically framed in the context of language evolution, and in relation to hypotheses surrounding the emergence of structure on a combinatorial level. Further, modality is extended as a concept which also affects the design of signal spaces in experimental work.

As a starting point for the chapter, table 3.1 lists the main physical differences between the spoken and signed modalities. The table is based on a table by Meier (2002), but I have added some differences not found in the original table which are explored in this thesis and elsewhere. Differences in green are from Meier (2002) (p.7). It is worth noting that these differences are often not independent of each other.

Oral articulators	**Sign articulators**
Primarily auditory signals	Visual signals
Largely hidden	Move in a transparent space
Coupled to respiration	Not coupled to respiration
Fast signal fading	Slower signal fading
Relatively small	Relatively massive
Oral articulators smaller dimensionality	Sign articulators larger dimensionality
Not paired	Paired
Harder to produce signal elements simultaneously	Easier to produce signal elements simultaneously
Harder to generate iconic signals	Easier to generate iconic signals

Table 3.1: The differences between speech (bound by the production capabilities of the vocal tract) and sign (bound by the production capabilities of the hands). Differences included in Meier (2002) (p.7) are presented in green.

3.2 Modality Effects and the Emergence of Combinatorial Structure

Meier (2002) states that duality of patterning is not affected by modality. However, evidence from newly emerging sign languages suggests otherwise. Al-Sayyid Bedouin Sign Language (ASBL) and Central Taurus Sign Language are newly emerging sign languages which has been shown to have no minimal pairs, and therefore are argued to have no phonological (combinatorial) patterning (Sandler et al., 2011; Caselli et al., 2014). Further, even fully developed sign languages have few true minimal pairs (Sandler, 1996). As a result of these observations, I am interested in specifically what effect modality has on the emergence of combinatorial structure. Here, I talk about some general properties of signed languages that make them different to spoken languages and the implications this has for the emergence of combinatorial structure before specifically moving to hypotheses discussed in the previous chapter.

3.2.1 Auditory or Visual Signals

Perhaps the most obvious difference between the language modalities is that speech is primarily auditory (though users do use some visual information to process signals), and sign language is visual. This difference directly links to Meier's (2002) observation that the oral articulators are largely hidden, while the sign articulators are completely visible. However, this observation relies on a visual perspective. Speech is not "hidden" in terms of auditory transmission, which is how it is primarily perceived. However, the oral articulators not being visible may cause an extra hurdle for infants acquiring language, where learning the mapping between tongue position and sound production is not as straight-forward as it might be with visible manual gestures. Whether a system is auditory or visible will have an affect on the types of noise which can interfere with the system, the level of rapidity of fading possible, the amount of iconicity possible (depending on whether the meanings to be communicated as visual or auditory).

3.2.2 Dissociability and Simultaneously Presented Information

There are several salient differences between combinatorial structure in signed and spoken language. In spoken languages features are largely indissociable. A phoneme must simultaneously have voicing and place and manner of articulation. To a certain extent, this is also true of signed languages: phonemes must have a place, hand-shape, movement etc. However, sign phonemes can also be produced simultaneously, creating signals that are dissociable. While co-articulation does exist in spoken language, it is much more prevalent in signed languages and many more forms of co-articulation exist. For example, co-articulation between hands, within hands or with facial expressions. The visual system is capable of perceiving many more simultaneously presented aspects than the auditory system can (Brentari, 2002; Sandler et al., 2011), which perhaps creates a higher threshold for visual systems before a signal-space gets crowded. This is something that will have knock-on effects that I discuss in the sections below.

3.2.3 Ease of Articulation and Phonotactics

As discussed, different modalities produce different phonemes. In both modalities, some phonemes are easier to produce than others, making certain phonemes likely to appear again and again across the world's languages regardless of their relatedness. For example, nearly all spoken languages have at least one back vowel, which usually tends to be rounded, and nearly all languages have at least one front vowel, which usually tends to be unrounded (Schwartz et al., 1997). There is some work emerging which explores how different physiological abilities across a population can affect languages within modalities. Moisik & Dediu (2017) investigated whether populations where languages use clicks (the Khoisan population in Southern Africa) had anatomical features which make clicks easier to produce (the lack of an alveolar ridge). There have also been more general studies looking at how spoken language has adapted to be easy to articulate based on the amount of biomechanical force involved (e.g. Kirchner, 1998). Work has also been done on ease of articulation in signed languages, showing that signs adapt to use fewer joints and joints further away from the torso, to reduce the amount of mass being moved (Napoli et al., 2014). Results such as these may explain to some extent why some phonemes are more likely to appear in both spoken and signed languages.

Further to the idea that some phonemes are easier to produce than others, it is also true that some sequences of phonemes are more likely to occur together within a modality (phonotactics). In both language modalities, some combinations of phonemes are more likely than others because of ease of articulation. For example, syllable finally, /mp/ is more common in English than /np/ because /m/ and /p/ both share the bilabial place of articulation, making it easier to pronounce. This is an example of a phonotactic constraint. One thing that might affect the syllable structure in spoken language is indeed the biomechanics of the mandible. MacNeilage & Davis (1993) argue that the oscillation of the mandible could have provided a frame for the production of syllables. That is, the opening and closing of the mouth could create a frame for consonant-vowel patterns we often see in spoken language. Meier (2002) argues that there is no comparable joint with similar patterns of oscillation in the manual modality.

In signed languages, there has been a lot less research done on phonotactics, but we do know that phonotactics operate in a different way, so instead of phonotactics referring to rules about likely sequences, they instead refer to rules about likely co-articulations. For example, one hand mirroring the hand shape and configuration of the other in symmetrical signs in ASL (Battison, 1978). Accordingly, phonotactics is something else that differs between the modalities that may have knock on effects on the nature of emerging combinatorial structure.

Ease of articulation can also affect the structure that emerges in artificial language experiments. With slide whistles (used in Verhoef et al., 2014) it is easier to produce certain patterns than others. For example, if someone has reached the lowest sound possible with the whistle, by pulling the plunger all the way out, it is then easier to generate a rising tone by pushing the plunger back in than to reset the plunger to another position. Ease of articulation is also likely to affect structure we see in experiments that use discrete signal spaces. For instance, in Cornish et al. (2013), participants play the Simon game, where they are to memorise a sequence of colours and then reproduce the sequence. These reproductions are taught to a new participant in a transmission chain. Over generations, the signals are found to become easier to learn. However, the colours are arranged in a circle, meaning that some signals may be easier to learn precisely because they are easy to produce (i.e. just going around in a circle or repeating the same colour). While I would not argue that issues of ease of articulation make the findings from these experiments redundant, I would argue that this is a modality effect which needs to be understood and considered in the interpretation of results from artificial language experiments.

3.3 Emergence Hypotheses and Modality

I have considered above some modality effects broadly. In this section I will specifically focus on the hypotheses listed in the previous chapter and how they relate to modality effects.

3.3.1 The Pressure of Learnability

There is no evidence to suggest that spoken or signed languages are any more or less learnable than the other. The rate of acquisition is the same for both (Newport & Meier, 1985). However, there is some evidence to suggest that iconic signs and systematic sound–meaning mappings are easier to learn at the early stages of acquisition than fully arbitrary ones (see Monaghan et al., 2014), and it may be true that signed languages have more iconicity in them than spoken languages, but this will be discussed in the iconicity section below.

3.3.2 Signal Space Crowding

Del Giudice (2012) hypothesised that the lack of phonological patterning in emerging signed languages is the result of the signal space in signed languages being much larger than in spoken languages. A bigger signal space will allow for a greater number of distinct signals without the need for combinatoriality. The signal space in signed languages is obviously very different than it is in spoken languages. Thinking of the articulators for spoken language only as the vocal tract, and the articulators of signed language as the arms, hands and upper body, it is uncontroversial to claim that the vocal tract is physically smaller in area. However, that does not paint the whole picture. While the production space is much smaller than the other, the perception space is not. To understand, think of the small movement your tongue needs to make to create the perceptual difference between [s] and [ʃ]. Now image the same amount of movement in your finger. It is barely perceptible, and it certainly would not help to differentiate phonemes in a signed language. So, we need to think about the size of the perceptual space, rather than the size of the production space. This is quite difficult to quantify, especially as one perceptual space is visual and the other is auditory. One thing we can do is look at the size of phonetic inventories. In spoken languages, phonemic inventories have, on average, around 34 phonemes (Maddieson & Disner, 1984). Phoneme counting is much more difficult in signed languages as true minimal pairs are much less frequent in signed languages (Sandler, 1996). Despite the difficulty in counting, many have argued that signed languages have a bigger phonological (signal) space.

Morgan (2016) argues that a large phonological space in signed languages is the result of words in signed languages not being composed of strings of building blocks (something also observed by Crasborn et al., 2009). However, the signal-space crowding hypotheses would argue that this relationship is in fact backwards, that the large phonological space is the thing that would allow for so many holistic forms to exist without combinatorial structure. Indeed, the fact that ASBL can exist with thousands of words without a level of combinatorial structure suggests that signal space crowding has little effect in signed languages. Morgan (2016) makes the argument, using data from Kenyan Sign Language, that instead of combining forms sequentially, signed languages create distinctions by having many simultaneous features. These features are things that can make meaningful distinctions between signals, or what I referred to in Chapter 2 as dimensions. Perhaps, it is then the dimensionality of a signal space which has the biggest effect on its size.

3.3.2.1 Signal Space Dimensionality

The spoken modality has several ways phonemes can differ. Place of articulation, manner of articulation and voicing are the main ones. These would be considered the dimensions in spoken language using the definition I outlined in the Chapter 2.

Crasborn et al. (2002) claim that signed languages have more dimensions in their phonemic spaces than spoken languages. They illustrate this with the following example: in spoken language it is not possible to separate the notions of place of articulation and the articulator used. For example, a bilabial phoneme must be produced using the lips, and phonemes produced in the oral cavity must be produced with the tongue. However, in the manual modality, the articulator (which they argue is the hand-shape) is distinct from the place of articulation, which provides an extra dimension in signal space of sign, before one even considers orientation and movement in the sign. There are also facial movements which can create distinctions between signs. It is because of all of these dimensions in sign languages that so many distinctions are possible. Morgan (2016) argues it is simultaneously presented features that negate the need for sequential signals, which is, in principle, the dimensionality of the signal space overcoming the problem of signal space crowding.

As we saw in Chapter 2, dimensionality also has an effect on the amount of iconicity possible with a signal space. If it is true that the manual modality has more dimensions, this may be at the root of why it is so much more iconic as a medium. I discuss this in the section on iconicity below.

3.3.2.2 Effects of Noise

All communication channels have noise, and all communication systems need to be robust against noise. However, different modalities will be subject to different types of noise and to different degrees of noise. In both signed and spoken languages, internal noise can come from the variability in signal production and signal perception. In spoken language, external noise comes in the form of audio interference from the environment, whether that is in the form of the challenges of trying to communicate over a larger than average distance, other people speaking simultaneously in a crowd, or a bad phone line. Signed languages are also subject to problems from visually noisy environments. Communicating over large distances presents the same problems as it does with spoken language. Many signers in one environment can be distracting, and noise can also come in the form of a bad connection on a video call.

Quantifying whether one modality is more subject to noise than the other is difficult, because noise comes in so many different forms. It may be intuitive to guess that spoken language has more noise, maybe because the word "noise" nearly always refers to an auditory phenomenon in English. However, it is possibly also true that it is easier to focus on one person's signal if they are signing compared to speaking because auditory sources are harder to localise. It is harder to not pay attention to other sources of sound in the environment than it is to ignore visual distractors that appear in a different place in the visual field. With signed languages, in a crowd, it is not possible to confuse one person's signal as coming from someone else, because you have to look at the person to perceive the signal. It is also relatively easy to distinguish a person signing from any amount of movement going on around them. However, with spoken language, it is possible for one person's voice to get lost in a sea of noise if too many people are speaking at once.

3.3.2.3 Nonlinearities in a signal space

I have already discussed how nonlinearity in the signal space relates to speech in the previous chapter. Janssen et al. (2016) use the example that fricatives produced along the hard palate are often very low in their perceptual variability, but once the tongue reaches the alveolar ridge, then there is a sharp perceptual change between [s] and [ʃ]. There is no research that I am aware of that looks at nonlinearity in the manual modality, though as the difference between production and perception in the manual modality is much more transparent than with the spoken modality, I would expect there to be less of a nonlinearity effect in the manual modality.

3.3.2.4 Effects of Signal Duration

It is possible that modality constrains how long a signal can be. Speech is constrained by breath, which is not something that the manual modality is constrained by. There may be modality effects related to production and processing effort. Klima & Bellugi (1980) demonstrate that the rate of production in American Sign Language (ASL) is much slower than it is in spoken English. Klima & Bellugi (1980) go on to argue that the slower rate at which signed languages can operate is one thing that might cause the simultaneous presentation of features. That is, because the time dimension does not allow for sequential signs, all this information is presented at once. This account relies on an assumption that a time constraint is at work, which is possible for all the reasons I mentioned in the last chapter. Further, many contexts for using signed languages in the modern, urban world involve the use of interpreters translating spoken language to signed language, creating a pressure for signed language to match the speed of spoken language.

Signal duration is also very relevant for modality effects in experimental work, where some experiments put a limit on the duration which signals can be, and others do not. These are discussed thoroughly in Chapter 9 of this thesis.

3.3.2.5 Rapidity of Fading

Rapidity of fading is something that is much more different between the spoken and manual modalities. Speech has a higher rate of rapidity of fading than the manual modality, or at least the rate of rapidity of fading in signed languages is a lot more flexible. Auditory signals necessarily have fast rapidity of fading while visual signals are a lot more flexible in this regard. Manual signs can last as long as the producer would like. Spoken signals can also be drawn out to a certain extent, which they sometimes are to create emphasis, e.g. saying "the long cat is looooooooong" to emphasise the length of a cat. However, the length of a spoken signal can only last as long as a person has breath, and stretching out phonemes can only occur on a subset of possible phonemes (it is more difficult, for example, to exaggerate the length of a plosive), and on one phoneme at a time.

If we do assume that speech has faster rapidity of fading than sign language, this may facilitate the emergence of combinatorial structure in speech to a greater extent than it might in the sign modality, following the results of Galantucci et al. (2010).

The effects of rapidity of fading are also something which should be considered as a modality effect in artificial language experiments. Some experiments have rapidity of fading, for example experiments with auditory signals (Verhoef et al., 2014), and other do not, for example experiments with graphical signals (Garrod et al., 2007). However, rapidity of fading (and signal duration for that matter) is rarely listed as an experimental design choice, just a side effect of the particular signalling apparatus being used. As Galantucci et al. (2010) has shown that it has an effect, then it should be considered as a factor in the design of every artificial signalling experiment.

3.3.3 Iconicity

Iconic signals are possible using both the manual and spoken modalities. However, recent experimental evidence has shown that it is easier to be iconic using the manual modality (Fay et al., 2014a). Further, it is generally accepted that there is more transparent iconicity present in signed languages than in spoken languages. This is difficult to measure, but even in signs that have become somewhat conventionalised, their iconic beginnings are often transparent in signed languages, for example, indicating a whisker to denote CAT in ASL.

I discussed above that the manual modality has more dimensions than the spoken modality, and this may be one thing that contributes to its ability to be iconic. The real world is a highly complex, multidimensional meaning space. When the number of signal dimensions matches the number of meaning dimensions, then iconicity is a lot easier because there is then the possibility for a one-to-one mapping between the spaces. This also works in terms of degree, the more similar the number of dimensions (or topology) between signal and meaning spaces, the easier it will be to create iconic signals. This means the signal spaces with higher dimensionality will be better suited for creating iconic signals which refer to the real world.

The idea that iconicity in the manual modality may inhibit the emergence of combinatorial structure is not a new one. Goldin-Meadow & McNeill (1999) made the argument that spoken languages rely so heavily on combinatorial structure because it is less suited to iconically represent meanings. The hypothesis follows that if an element of a signal loses iconicity, or never had it, then it becomes available to be reanalysed as a meaningless unit which can be recombined. Sandler (1996) also argue that the lack of true minimal pairs in signed languages may be because of the iconic foundations of signs: iconicity may lead to phonologically irrelevant information persisting in signs.

Another angle is to not think about iconicity as something which is inhibiting re-analysis of signal-elements, but as facilitating large numbers of holistic vocabulary items to exist in the first place. Large numbers of holistic signals may be more difficult to learn without the presence of iconicity.

3.4 Discussion

This chapter has collated hypotheses and empirical evidence for how the differences between the manual modality and the spoken modality affect the emergence of combinatorial structure. It is relatively easy to observe differences between the modalities. However, it is difficult to isolate cause and effect in natural data because there are so many factors. Further, data from emerging signed languages is difficult to compare with any known spoken language. Most "young" spoken languages are pidgins or creoles that already have existing languages as their starting point. In order to tackle this problem, this thesis aims to isolate the physical differences between the manual and spoken modalities in order to see the effects of these differences when isolated in controlled experiments.

In Chapter 2, there was a discussion on whether combinatorial or compositional structure came first. The current chapter has discussed further evidence from emerging sign languages that has shown a delayed emergence of combinatorial structure, which provides evidence that compositional systems can predate the emergence of combinatorial structure. However, what this chapter has argued, is that the factors at work that might cause that delay in the emergence of combinatorial structure may be different or smaller in the spoken modality. Therefore, I propose that the question of which came first is dependent on the modality which is being used to produce signals. I will revisit this claim again in the conclusion.

CHAPTER 4

Experimental Methods [1]

This chapter will cover the experimental methods used throughout the thesis. The chapter discusses the different structures of cultural learning experiments used in evolutionary linguistics, and how they relate to evolutionary pressures in the real world. I then go on to discuss continuous signal space proxies used in previous experimental work on the emergence of combinatorial structure, including slide whistles. The discussion on previous methods will cover the advantages and disadvantages of using the various signal space proxies in order to put the methodological choices in the current work in context. The chapter closes with a discussion of the main signal space proxy used in this thesis (in Chapters 6-9), the Leap Motion. The chapter also goes into the structure of the Leap Motion experiments presented throughout the thesis in order to not have to repeat the paradigm and methods in every chapter. Experiment 1 is the only experiment whose structure is not covered in this chapter because it does not use the Leap Motion and so is "the odd one out"; to prevent too much flipping backwards and forwards for the reader, its methods and procedure are isolated to its own chapter.

[1]Parts of this chapter are published in Behavior Research Methods (Eryılmaz & Little, 2016). Parts of this chapter pertaining to signal measures are also included in Cognition (Little et al., 2017b)

4.1 Artificial Language Experiments

Artificial language experiments[2] are one way we can simulate cultural learning in the lab. Currently, they are one of the primary methods used in the field of evolutionary linguistics (see Scott-Phillips & Kirby, 2010, for a review). Artificial language experiments originated in linguistics and psychology (originally by Esper, 1925) as a method to isolate specific linguistic features, in order to understand how they are learnt and used without interference from existing linguistic knowledge. In the past decade, these experiments have been used in experimental set-ups that model the cultural evolution of language, either via transmission (e.g. Kirby et al., 2008), communication (e.g. Selten & Warglien, 2007) or both (e.g. Kirby et al., 2015).

Cultural learning experiments were originally inspired as a way to validate computational models investigating the evolution of language on a population scale. Previously, one of the main criticisms of computational models was that they may lack ecological validity. Artificial language experiments are a good way to provide computational work with the validity it is lacking. These experiments make it possible to observe how individual-level biases accumulate to population-wide effects within real subjects. On the other hand, modelling work still has its strength in the scale of the numbers it can deal with.

In this section, I will briefly cover the 3 types of artificial language experiments used in this thesis (iterated learning, social coordination and individual signal creation), and how they relate to processes in the real world.

4.1.1 Iterated Learning Experiments

The iterated learning methodology was first used as a way to model cultural transmission and accumulation by Bartlett (1933). However, the paradigm was co-opted for work investigating the evolution of language much more recently by Kirby et al. (2008). This work was directly inspired by agent-based modelling work that uses iterated learning to look at how linguistic structure emerges and changes (e.g. Griffiths & Kalish, 2007; Kirby et al., 2004). Iterated learning experiments model the vertical transmission of language and how learning biases can be amplified by repeated

[2]While in the literature at large such experiments are called Artificial Language Learning experiments (ALL), in this thesis, I have contracted this to "Artificial Language Experiments" because not all of (or even most) of the experiments presented here have their focus on learning, though they still utilise artificial languages.

learning and transmission over generations.

Artificial languages used in iterated learning experiments usually start off as small, unstructured, holistic signals that are paired to meanings. The signal-meaning pairs are taught to a participant who is then asked to reproduce the signals from memory, usually by labelling meanings. Their output is then recorded to be taught to the next participant (or next 'generation') in a chain. This process gets repeated for as many generations as the experimenter would like (usually between 5 and 10). The outcome of this process is typically the emergence of structure in signals that reflects the structure in the meaning space (as happened in Kirby et al., 2008). However, not all studies use meanings (e.g. Verhoef et al., 2011). As discussed in Chapter 2, most of the signals used in iterated learning experiments start with a set of discrete building blocks (typically letters arranged into syllables). However, these experimental semiotic studies have started to use continuous signal spaces more and more (e.g. Verhoef, 2012;Verhoef et al., 2015, and in Chapter 5 of this thesis in Experiment 1). As the primary experimental tool used in this thesis, these signal spaces are reviewed in this chapter.

4.1.1.1 Learning Bottlenecks

Learning bottlenecks in iterated learning experiments (as covered in section 2.4.1) can either be artificially imposed, as they were in Kirby et al. (2008), with participants only seeing a set number of signal-meaning pairs, but being asked to create signals for meanings they have not previously seen. However, learning bottlenecks can also arise as the result of the task being too difficult to complete perfectly. For example, in Verhoef (2012), participants heard all of the signals they were asked to reproduce, but most participants found that they could not remember all signals perfectly, meaning that they had to extrapolate knowledge of signals they did remember to make up for those they had forgotten.

4.1.2 Social Coordination Experiments

While iterated learning experiments model the vertical transmission of language, social coordination experiments model the horizontal transmission of language. These experiments most commonly come in the form of communication games where two or more participants are not given an initial starting language, but must bootstrap a communication system themselves via interaction with each other (as in Experiment 6 in this thesis). In these experiments, participants can either be assigned the roles of "speaker" and "hearer", or they can take it in turns to take on these roles. Experiments can include interaction, opportunities for repair and feedback, but do not have to. Communication games can also be used in iterated learning experiments, where the output from one pair of participants is taught to a new pair (for example in Carr et al., 2016; Kirby et al., 2015; Janssen et al., 2016; Verhoef et al., 2015). There have also been studies that compare the effects of iterated learning on its own with communication on its own (Carr et al., 2016; Fay et al., 2010).

4.1.3 Signal Creation Experiments

In this thesis, the majority of experiments are signal creation experiments (Experiments 2, 3, 4, 5 and 8), which is a paradigm I designed. The paradigm is modelled on a social coordination experiment. However, instead of having multiple participants communicating with each other, these experiments just have one participant who communicates with themselves. The exact dynamics of this process are detailed later in the chapter. In these experiments, participants need to recognise the signals they have generated, which creates a pressure for expressivity in the experiment (the need to have signals be different from one another), without the need to have more than one participant in the lab at a time. This method makes experiments easier to run, especially when participant recruitment is a problem. However, there are some issues regarding participants responding differently if there is another participant interacting with them. For example, signals produced in individual signal creation experiments will not be subject to pressures of transparency and informativeness in the same way that they are in communication experiments. This means that while these experiments might be able to tell us about cognitive processes that result in the initial production of linguistic systems, they are probably more limited in what they can say about the evolution of language after a very initial stage. However, they still allow me compare behaviour between conditions where modalities or signal-meaning mappings are different, and these differences should still be valid when interaction is brought into the equation. I discuss these issues at further length in Chapter 8, where I compare a communication game with 2 participants with one of these signal creation experiments, and in the conclusion of the thesis under the further work section.

4.2 Continuous Artificial Signal Spaces

As I outlined above, artificial language experiments allow us to investigate linguistic features without pre-existing linguistic knowledge interfering. Creating an artificial language helps to eliminate interference from pre-existing linguistic knowledge. However, if signals are still very language-like (e.g. they are made up from phonemes or letters), then existing linguistic knowledge is still likely to get in the way of participants' behaviour. As a result of this, an increasing body of work is appearing that uses continuous signal spaces in order that discrete elements can emerge. Below, I discuss previous experimental work that uses continuous signal spaces, some of which has already been discussed in Chapter, 2 but specifically focusing on the benefits and shortcomings of the various paradigms.

Further, I explain how the experiments in this thesis, which use the Leap Motion, deal with these issues. Specifically, I discuss the problem of interference from iconicity, the problem of data that is difficult or labour intensive to analyse, and the problem of having restrictions on the shape and size of artificial signal spaces. This breakdown will hopefully help illustrate how the experiments in this thesis improve or build on previous work.

4.2.1 Limiting Opportunity for Iconicity

The use of graphical signals to investigate trends in how communication systems emerge and evolve started with the use of graphical symbols by Healey et al. (2002), and has since grown into its own field of research: "experimental semiotics" (for a review see Galantucci & Garrod, 2011; Galantucci et al., 2012). Experimental Semiotics has significant overlap with artificial language experiments in evolutionary linguistics, and is not confined to only graphical signals. However, many of these experiments investigate the effects of communication and transmission by using graphical pictionary-style communication tasks where participants are given a concept to communicate without the use of words (e.g. Garrod et al., 2007; Fay et al., 2008). These experiments are useful for investigating processes such as conventionalisation. However, experiments investigating the emergence of combinatorial structure become difficult to design with graphical paradigms, as participants are very familiar with presenting content graphically, both using written language and drawing. Graphical interfaces make it easy to utilise iconicity, which has been shown to affect the emergence of combinatorial structure (see for instance Roberts et al., 2015 and Verhoef et al., 2015). Further, different levels of iconicity are possible using different language modalities (Fay et al., 2014b), meaning that trends demonstrated with paradigms offering a lot of available iconicity (drawing) may not be extrapolatable to communication mediums with less available iconicity (e.g. speech).

In order to combat these issues of iconicity, Galantucci (2005) developed an approach that used a graphical interface, but had constraints on what participants could do using the apparatus. The interface consists of a stylus that writes on virtual paper which constantly drifts downwards, so participants can only control signals on the horizontal dimension. Using this approach, Galantucci has conducted social coordination experiments to investigate how communication systems with combinatorial structure can emerge (Galantucci, 2005; Roberts & Galantucci, 2012). More specific experiments have looked at how rapidity of fading (how quickly signals disappear after production) affects combinatorial structure in signals (Galantucci et al., 2010), and how the potential for iconicity affects the emergence of combinatorial structure (Roberts et al., 2015). There have also been experiments that have used the apparatus in an iterated learning paradigm, where participants' outputs were fed to other participants in transmission chains to investigate whether combinatorial structure emerges more reliably via vertical transmission (learning) in contrast with horizontal transmission (communication) (Del Giudice, 2012).

4.2.2 Ease of Analysis

Verhoef has done several experiments that use slide whistles as a proxy for an articulation space (e.g. Verhoef et al., 2014). Her experiments involve participants learning pre-recorded whistles and reproducing them from memory. This has been implemented in an iterated learning experiment to see if individual learning biases within transmission chains could influence the emergence of combinatorial structure in an inventory of whistles. Verhoef has implemented this in conditions with meanings to investigate the role of iconicity (Verhoef et al., 2015a) and without meanings, to isolate only the role of learning biases (Verhoef et al., 2014). Results have clearly shown that within transmission chains, combinatorial structure emerges and signals become more learnable. However, the output from these experiments are audio recordings of the acoustic signals, meaning that quite a bit of processing is required in order to extract the relevant data (the pitch values at each time frame) before analysis can start. Further, the relationship between stopper-movement and signal-pitch is not linear, and manipulating the mapping between stopper and pitch is not possible. This not only restricts possible experimental designs, but also complicates analysis if the experimenter wishes to calculate something like the stopper position from the pitch data.

Since these initial experiments, a computational alternative to the slide whistle has been developed that can work with a mouse on a screen, or via touch pads on tablets. In one experiment, participants created signals by placing their finger on a virtual slide whistle app on a tablet (Verhoef et al., 2015b). This experiment explored whether some meanings being more easily mappable than others would facilitate communicative success. This has undoubtedly solved the problem of having data which can be analysed quickly without much processing[3] Verhoef has since used a digital signalling apparatus but without auditory feedback, simply having visual signals represented by a bubble which can be moved up and down using a touchscreen (Verhoef et al., 2016).

Both Galantucci (2005) and Verhoef's slide whistle experiments have used gaps in the signals to measure structure. The analysis (e.g. in Roberts et al., 2015) starts with a signal already segmented into "forms" where signal parts are separated by a gap (e.g. the stylus lifting off the pad, or the participant stops blowing into the whistle). In other words, the participants were allowed to leave gaps at segment boundaries. These gaps are problematic because they are pre-conventionalised markers for structure. Instead of relying on statistical regularities or negotiating a boundary marker themselves, they are given an explicit way to segment the signals. Comparable cues are available in written language (such as spaces between words), but not necessarily in human speech, and certainly not at the phonemic level.

4.2.3 Flexibility of the Signal Space

Using slide whistles, the signal space is difficult to manipulate. An attempt has been made to change the size of the signal space using slide whistles by putting a stopper on the plunger in this thesis (Chapter 5).

Using the stylus paradigm (Galantucci, 2005), it is also difficult to manipulate the size, shape or dynamics of the signal space because it is confined to one dimension, causing researchers to manipulate the meaning space, rather than the signal space, to

[3]In Verhoef et al. (2015b) the researchers analysed some of the data as it was being produced in real time, which was especially useful because some of the data was collected as part of a public exhibition so it was nice for participants or visitors to be able to see the data being produced and analysed in real time.

investigate the effect of things such as iconicity (e.g. in Roberts et al., 2015).

Using digital slide whistles, it is easier to manipulate the shape of a signal space, and an experiment has been done which looks at the effects of different biases created by nonlinear mappings between the signal-space and the auditory feedback (Janssen et al., 2016). However, being on a flat surface, participants are still often tempted to produce signals as if they were graphical, focusing on the articulation space, rather than the auditory aspects of the signal. Because of this, the digital paradigms mentioned here have so far been limited to a 1 dimensional space (usually pitch). With a 2 dimensional space on a flat surface, participants are even more tempted to just "draw" their referent, as happened in de Boer & Verhoef (2012).

4.3 Leap Motion

The proxy used in the majority of this thesis (chapters 6-9) is the "Leap Motion" (Holz, 2014). The leap motion is an infrared sensor that detects hand placement and manual movement in the 3D space above it. With technical assistance from fellow PhD student, Kerem Eryılmaz, I designed a signal-creating interface that gives audio feedback based on where a participant's hand is in relation to the device. The signals produced are similar to signals produced with the musical instrument, the theremin. A participant can produce an auditory tone by placing their hand above the sensor, and then affect the pitch or volume of the tone by moving their hand. The exact dynamics of this are explained in more detail later in the chapter.

4.3.1 Advantages of the Leap Motion approach

I set out specifically to design a framework that limits opportunities for iconicity, improves the ease of analysis and, most importantly, is flexible in its geometry, size and in the nature of the signals it can produce in order that modality effects can be experimentally investigated.

4.3.1.1 Limits Opportunity for Iconicity

The Leap Motion framework generates auditory signals which are very restrictive, affording less iconicity than modalities that produce graphic signals. The framework still has a visual element (i.e. the hand position in front of the participant) and, as a result, participants still use this information to try and generate visually iconic signals. However, it is the auditory signals that are transmitted, not the visual ones, which makes iconicity a less salient feature in the transmitted signals. Iconicity in signals could be absolutely prevented by transforming mappings between hand-position and auditory feedback, or by designing the signal space to be less intuitive with a given meaning space. Importantly, the framework offers flexibility to make the opportunity for iconicity more or less possible.

4.3.1.2 Ease of Analysis

Using digital sensors such as Leap Motion, raw, numeric experimental data is available as soon as the experiment is finished in neat data frames.

4.3.1.3 Flexible Signal Space

It is possible to change the shape and dimensionality of the signal space using the Leap Motion framework. Having a flexible signal space makes it easier to make a signal space more or less like a signal modality used in natural languages. For example, slide whistles are more constrained than speech, but the results from these experiments are extrapolated to be relevant to speech. The current framework allows for the signal space to be more or less like speech, or more or less like gesture, in order to answer whether previous signal spaces have ecological validity. Further, it allows for comparison of data from signal spaces that differ in a feature, perhaps relevant to the differences between the spoken and manual modalities, or differences between previously used artificial signal spaces.

4.3.2 Signals

Using the Leap Motion, auditory signals are produced by moving one hand above the device. In the experiments in this thesis, the signals can be manipulated along the horizontal axis (which manipulates pitch low to high from left to right), or the vertical axis (which manipulates volume quiet to loud from up to down). Not all experiments use both. The paradigm records hand position above the Leap Motion in three spatial dimensions (though, in this thesis, only 2 at a maximum are used) and the duration of signals is also measured.

Auditory feedback was calculated from the x- and y-coordinates of the centre of participant's hand (see figure 4.1). The amplitude, manipulated on the y-axis, and fre-

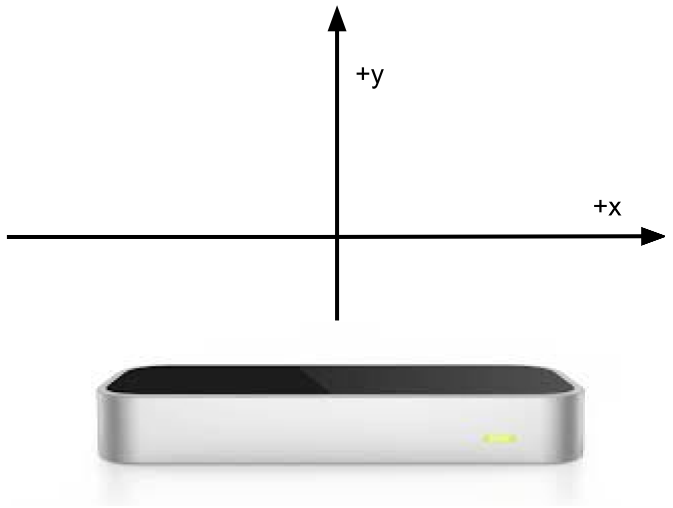

Figure 4.1: The signal dimensions available in phases with a two dimensional meaning spaces (relevant to experiments in Chapters 6 and 7.

quency, manipulated on the x-axis, was non-linear in relation to hand-position. Transformations that produced the auditory feedback were logarithmic and exponential respectively (see equations below). The feedback was coded in this way because of the way humans perceive both pitch and volume. When the mapping was linear, participants in pilots found it very difficult to differentiate tones with high frequencies, or with low amplitudes.

$$amplitude = 1.1 - \frac{\log|y|}{\log 250}$$

$$frequency = 110 \times 3^{\frac{(|x+200|)}{200}}$$

The articulation space of the leap motion was $\sim$500 coordinates across and $\sim$500 coordinates high (coordinates were 1mm apart). There was some scope to produce signals beyond this space, and no need to produce signals that filled the space. Accordingly, there can be quite some individual differences between the ranges of coordinates used in signals, but when the variation of signals was used as a measure in the analysis, I always controlled for participant number.

When a participant's hand was taken out of the signal-space, the auditory feedback stopped. Pauses or gaps created in this way were not recorded, so if a pause was generated in production, it was deleted when the signal got played back. I was keen to not have pauses in the signals in order to make the signals more analogous to speech and not give participants a way to segment the signals themselves (as they can in Galantucci, 2005, and Verhoef et al., 2014). When a participant's hand is withdrawn during an experiment, the signal fades, rather than cutting off immediately, which made the signals sound more continuous.

Signals throughout the experiments in this thesis had no time limit, except in Experiment 8, which explicitly tests the effects of time pressures of signal structure.

4.3.3 Ecological validity of signals

The Leap Motion uses hand movement as its input, which can potentially thought of as gesture. This raises the problem that participants' knowledge of gesture could influence the signals they make. While it is true that gesture is used in communication almost ubiquitously, the gesture that produces signals using this framework is not similar. For one, precise placement of the palm of one hand is not an important feature of co-speech gesture or gesture in sign language. Further, the use of hand-placement to generate precise auditory feedback is not something that occurs in natural language.

When humans acquire spoken language, they must work out how the movements of their tongue and vocal tract map to the sounds they can produce themselves, and the sounds being produced by those around them. In Chapter 3, I wrote briefly about how this task may be more tricky in the spoken modality because the articulators are visually hidden, while they are not in the sign modality. While the articulators used to produce signals with the leap motion (the hand) is visible, participants still have to overcome the problem of being able to map physical movement to auditory feedback, perhaps making it more similar to speech.

4.4 Structure of Signal Creation Experiments

The majority of experiments in this thesis are individual signal creation experiments (Experiments 2, 3, 4, 5, 8). These experiments are modelled on social coordination experiments, but only one participant is involved in each experiment. Experiments are made up from signal creation tasks, where participants are presented with meanings and asked to create signals for each one, and signal recognition tasks, where they hear their own signals and are given a forced choice task between a set of signals.

4.4.1 Signal creation tasks

During a signal creation task, participants see a meaning for which they must create a signal. The signal creation screen (see Figure 4.2) features the image, and a "Record" button that can be pressed to start recording a signal. This turns into a “Stop” button which can be pressed to stop the recording when the participant is done creating their signal. After the participant has created a signal, the button turns into a “Rerecord” button for if the participant is unhappy with their first recording. Participants can also play back a signal they have just created by pressing “Play”. When participants are happy with their signal they can then press “Submit”.

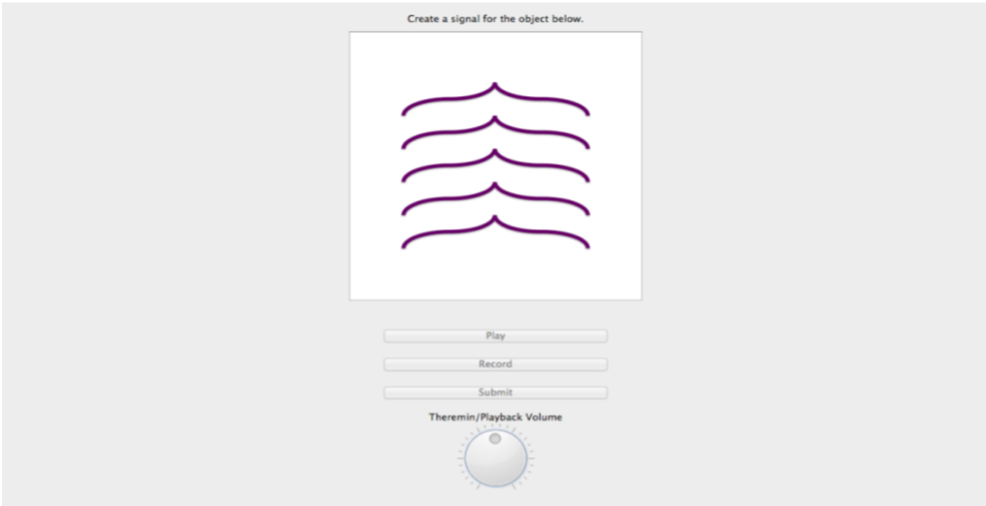

Figure 4.2: A signal creation screen. The meaning is taken from the set of meanings used in Experiments 5-8.

4.4.2 Signal recognition tasks

The signal recognition task starts with instructions to the effect of: “choose the image you think the signal refers to” (see Figure 4.3). There is a “Play” button that the participant can press to hear the signal to be recognised. There is also a set of possible meanings that includes the target meaning that participants can select and deselect (only one is allowed to be selected at a time). There is a “Submit” button to confirm their selection. The participant is unable to press the submission button without listening to the signal at least once, preventing participants from guessing at random without knowledge of the signal.

4.4.3 Structure of Communication Game Experiments

There is only one communication game experiment in this thesis (Experiment 6), which uses essentially the same experimental setup as the individual signal creation

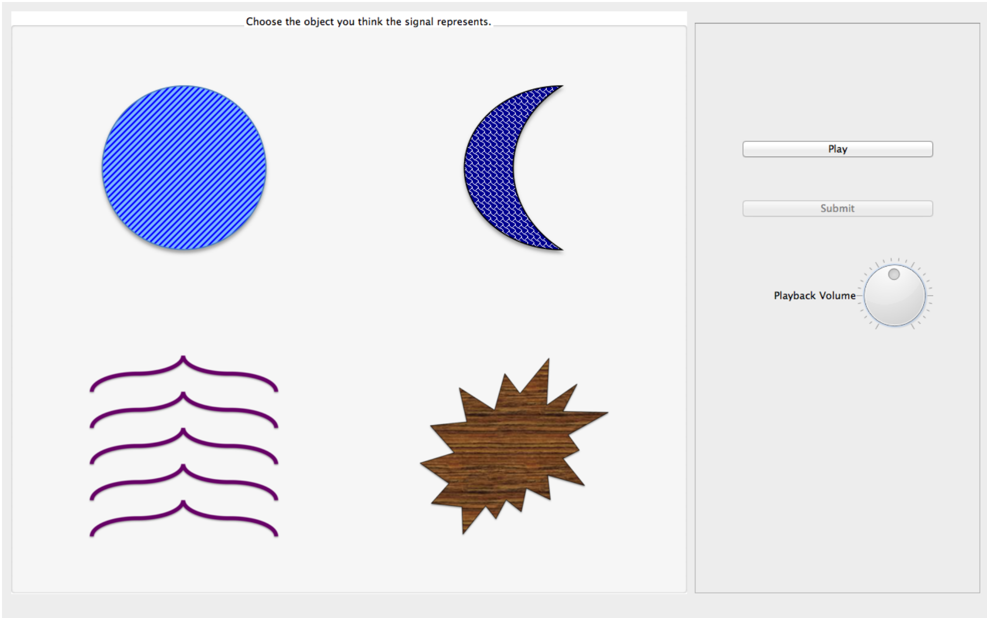

Figure 4.3: Signal recognition screen. The meanings here are from Experiments 5-8.

experiments, but with 2 participants instead of 1. The signal creation and recognition tasks are nearly identical to the individual experiments, but participants create signals for the other one to recognise. More explicit description of the procedure for the communication game in the thesis is included in Chapter 8.

4.5 Analysing Data

4.5.1 Description of Data output by the Leap Motion

The data produced by each experiment was 2 csv files. One contained the data from signals, and one contained data from responses.

The signal data set kept track of who had produced the signal, what phase they were in, what meaning the signal was for, and what the coordinates of the signal were at each data frame. This resulted in many hundreds of lines of data for each signal, as each signal was made up from several hundred data frames (approximately 110 data frames per second).

The response data set included information on which participant was responding, what phase they were in, what meaning the signal they were responding to was for and what meaning they chose.

4.5.2 Descriptive statistics for signals

As a starting point, I measure the properties of the signals produced using the Leap Motion using basic descriptive statistics. Each signal was represented as a list of coordinates that could be used to calculate the duration of the signal (how many coordinates in a signal) and mean pitch or volume (from the mean number on the relevant dimension).

4.5.3 Measuring structure

With continuous non-discretised data, measuring structure is very difficult. In previous studies, measures of structure have included manual discretisation and bigram counts or hierarchical clustering. Systematicity can also be measured in a signal repertoire, using measures such as compression or entropy (Verhoef et al., 2014). In this thesis, I use measures for the amount of movement in signals. Below, I briefly discuss the measures myself and colleagues have used for measuring structure in the experiments in this thesis, including the standard deviation of signal trajectories, the amount of predictability in signals (similar to measure for compression) and Hidden Markov Models.

4.5.3.1 Standard Deviations

I use standard deviations of signal trajectories throughout the thesis. The values used to discern the standard deviations were the x- or y- coordinates in each signal, which are described above. The standard deviation was just calculated by calculating the variance of the coordinate numbers in the list or coordinates in a signal. So the more of the signal space a participant used (or the more movement in a signal), the bigger the standard deviation.

4.5.3.2 Probability Measures

Throughout the thesis, one measure of "complexity" and structure is the predictability of a signal, given the rest of the signals in a repertoire. This measure can pick up on things such as repeated patterns within a signal repertoire, where patterns can make signals more predictable given the rest of a repertoire. If signals have very little variation within and between signals across a repertoire, the predictability score will be very high. This measure reliably correlates with measures such as compression and entropy.

The procedure to calculate the predictability is as follows:

1. Use the k-means algorithm to compute a set of clusters S of hand coordinates using the whole repertoire, which will be used to reduce the continuous-valued trajectories to discrete ones ($k = 150$).

2. Calculate the bigram probability distribution P for each symbol $x_i \in S$.

3. Use the bigram probabilities to calculate the negative log probability of each trajectory.

The choice of k was set quite high at 150 to ensure the quantisation was sufficiently fine grained. I plotted each signal repertoire and compared it with the clustered repertoire to ensure that any structure had been preserved after the clustering (see figure 4.4).

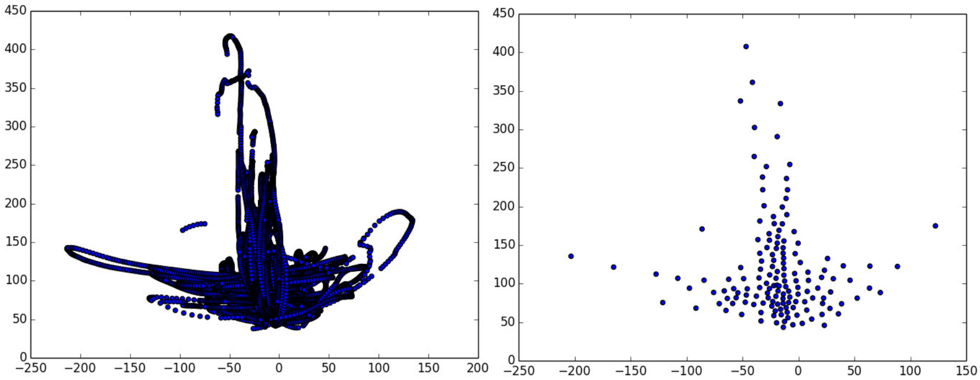

Figure 4.4: On the left is all signal trajectories from one participant mapped on to one another. On the right is the result of k-means clustering.

4.5.3.3 Hidden Markov Models

One method that has been used to measure structure in the data produced from the experiments presented throughout this thesis uses Hidden Markov Models (HMMs). HMMs are statistical models that assume the signals produced are Markovian processes with hidden states. Kerem Eryılmaz has developed HMMs as a way to model structure in the signals produced in my experiments by finding states that correspond to building blocks needed in signals for them to be fit for communication. These models aim to discover the phonological structure present in signals by being trained on a whole repertoire of signals from one participant (Eryılmaz et al., 2016). This work is the primary focus of Kerem Eryılmaz's PhD project "Markovian Models of Structure in Continuous Artificial Language Signals". Accordingly, a detailed methodology is not included in the current work, as the work is not mine (though the information that should allow replication is included in is presented in Little et al., 2017). However, I will include the results from this analysis work for Experiment 2 in the thesis, as I feel it does add to the understanding of the results.

4.5.4 Measuring Iconicity

There are two ways used to measure iconicity in this thesis. The first is designed to measure diagrammatic iconicity by objectively measuring which signals are more similar to one another and seeing if this corresponds to the structure in the meaning space. The other measure is a more subjective measure which can also account for more abstract or imagic forms of iconicity by having naïve participants listen to signals and asking them which meaning the signal refers to.

4.5.4.1 Measuring Diagrammatic Iconicity

In Experiments 2 and 3 measuring diagrammatic iconicity quantitatively was possible using simple descriptive statistics, as both the meaning space and signal repertoires were very structured and predictable. I go into more detail about this analysis in Chapter 6. However, when the meaning space was less structured (as in Experiment 4) then such straightforward mappings were not possible to draw. To combat this, Kerem Eryılmaz came up with a similarity measure based on HMMs which is summarised in Little et al. (2015). These similarity measures could then be correlated with the similarities between meanings.

4.5.4.2 Measuring Imagic Iconicity

In later experiments in the thesis (Experiments 5, 6 and 8) the meaning space was designed to be unstructured, making such objective methods for measuring iconicity more difficult. As such, I needed to employ more subjective measures. For Experiments 5 and 6, naïve participants listened to signals and paired them with meanings from an array of possibilities (Experiment 7). Various studies in the experimental semiotics literature have already done this (e.g. Garrod et al., 2007; Perlman et al. Roberts et al., 2015). If participants can pair signals with their intended meanings without any knowledge of how they were established, then those signals can be said to have imagic iconicity. I conducted these playback experiments online to allow for a massive number of participants, which increases statistical power.

4.6 Conclusion

In this chapter, I have outlined the different experimental paradigms used in this thesis. I have given a brief review of previous continuous signal spaces used in artificial language experiments. I have included details of the main signalling apparatus used throughout the thesis, detailing the experimental setups. I have further given detail on the analysis of data that will occur throughout the thesis, which can be referred back to when reading the results throughout. I have done this here as to not keep repeating the details, both of the experimental procedures and analysis, in every experimental chapter hereafter.

CHAPTER 5

Size of signalling space[1]

This chapter presents an iterated learning experiment using slide whistles. The experiment is designed to test the signal space size hypothesis outlined in chapter 2.

5.1 Introduction

This experiment has its primary focus in the size of the signal space. The effects of signal-space crowding have been discussed in Chapter 2. The signal space is not manipulated in its dimensionality, noise, transmission channel, or any other features, other than the amount of space inside the signal space, in this instance, a slide whistle.

Verhoef (2012) originally used slide whistles in an iterated learning experiment to explore the role of cognitive learning mechanisms in the emergence of combinatorial structure. Participants learnt signals that were unattached to meanings and then reproduced them. Their reproductions were taught to the next participant and so on. She found that the signals became more learnable as the result of the transmission process. Repeated elements also emerged within signals across a repertoire, which was cited as evidence for combinatorial structure.

The current experiment is a replication of Verhoef (2012). Meanings were not included in the current experiment in order to make it as close to the original Verhoef experiment as possible. Meanings were omitted in the original experiment because the authors were interested in the emergence of combinatorial rather than compositional

[1] Part of this chapter is published in the proceedings of EvoLangX (Little & de Boer, 2014).

structure. From previous experimental work by others (e.g. Kirby et al., 2008) and her own pilots, Verhoef (2012) knew that structure was likely to emerge that mirrored the structure in the meaning space, either with iconicity or compositional structure or both. In the absence of meanings, the structure that emerged in the original experiment is analogous to "bare phonology" which I discussed briefly in chapter 2. I write a little more about the methodological choice of omitting meaning and its implications in the discussion section of this chapter.

As I made clear in Chapter 2, I am not conducting this experiment in order to contradict the claims of Verhoef (2012). I do not dispute that learnability is a factor in the emergence of combinatorial structure. Indeed, the fact that combinatorial structure emerged in Verhoef's experiment with as little as 12 signals indicates that learnability was a stronger driver in that experiment than signal space size. However, with no comparison of a signal space with smaller (or bigger) size we can not know the exact effects of signal space size. As a result, I conducted the experiment presented in this chapter.

5.2 Hypothesis

The hypothesis investigated in this chapter is: does the size of the signal space cause signal space crowding and affect the emergence of combinatorial structure?

5.3 Experiment 1 Methods

5.3.1 Participants

48 people participated in 6 transmission chains, 8 participants in each chain. Participants had a mean age of 25.6 (s.d. = 7.6) with 22 male participants and 26 female. Participants were either recruited at the Vrije Universiteit Brussel, or the Humboltd Univeristät Berlin at a summer school for language evolution. Participants were paid €10 for the 1 hour (maximum) it took to complete the experiment.

5.3.2 Signals

Signals were produced using slide whistles produced by Grover-Trophy (also known as a swanny whistle, see figure 5.1). These were the same slide whistles used in the original Verhoef (2012) study. The plunger can be moved in and out in a continuous fashion in order to change the pitch of signals (the range of an unaltered whistle was between 450-2500Hz). The further out the plunger, the larger the chamber inside the whistle, and therefore the lower the sound. It is also possible to blow harder on the whistle in order to get a sound that is an octave higher.

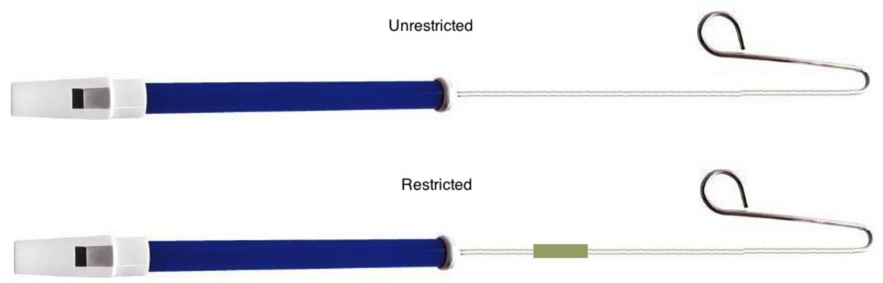

Figure 5.1: The conditions in the experiment, showing the stopper that restricted the pitch range in the restricted condition.

5.3.3 Conditions

There were 2 conditions:

1. **Unrestricted Condition**. This condition was a straightforward replication of Verhoef (2012), which had participants replicate whistle signals using the same normal, unaltered slide whistles used in the original experiment with pitches possible between 450-900Hz at normal blowing intensity.

2. **Restricted Condition**. This condition had the whistles altered by restricting the amount of space that the plunger could operate in. There was a stopper fixed on the plunger such that the plunger could not be extended by more than 4cm (see figure 5.1), this meant possible whistles could be between 450-600Hz at normal blowing intensity. As a result, there was a restricted amount of the signal space at the low end of the pitch range that could be used to create signals.

3 transmission chains were run with the unrestricted whistle. 3 transmission chains were ran using the redistricted whistle.

5.3.4 Procedure

Participants were given written and spoken instructions. All participants gave their informed consent for their signals to be recorded. All participants gave personal information, which was anonymised, about their age, gender, linguistic background and musical ability.

The experiment used the exact same experimental interface as Verhoef (2012). The code for the interface was kindly provided by Verhoef. The experiment is framed to participants as an alien language class. A little alien (see figure 5.3) greets the participant and explains the slide whistle as a way that the participant can replicate the alien sounds. These alien sounds are signals produced using a slide whistle which were either the initial set of signals for participant 1 in a chain (see section 5.3.5) or they were the signals produced by the participant before them in a chain (see figure 5.2).

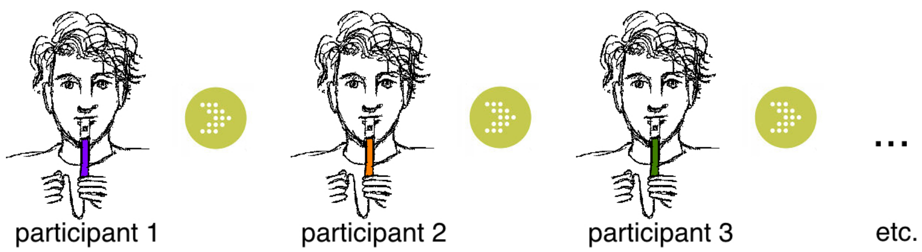

Figure 5.2: The output signals from one participant were used as the input for the next participant.

Each experiment started with a practice phase to get used to using the whistle, then there were 4 rounds of learning and recall phases. These phases will be explained below.

5.3.4.1 Practice Phase

The participants started with a practice round that involved imitating examples of bird songs. This phase existed so that participants could get used to how the whistle works. They were explicitly told that the sounds are bird songs, so they did not misinterpret this task as being part of the alien language. The practice phase was not recorded.

5.3.4.2 Learning Phase

After the practice phase, participants listened to 12 whistles one by one. They saw the screen shown in figure 5.3, which showed them how many of the 12 signals they had listened to and imitated. After listening to each signal they were asked to record themselves reproducing the whistle. All recordings were either done in a sound proof booth (in Brussels), or in an empty classroom (in Berlin).

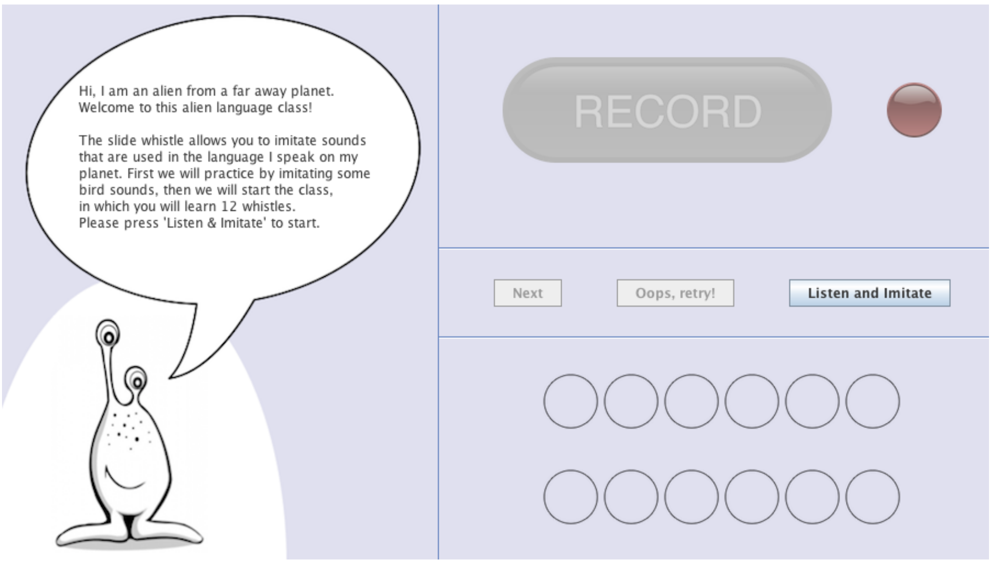

Figure 5.3: The interface participants saw in the learning phase of the experiment designed and coded by Verhoef (2012).

5.3.4.3 Recall Phase

After the learning phase, participants were asked to recall as many whistles from the learning task as they could. They were asked to produce 12 whistles in any order. If they could not remember whistles, then participants were explicitly told to produce whistles that they think might "fit well in the language". These were exactly the same instructions used in Verhoef (2012). Signals produced in the last recall phase participants completed in an experiment (the fourth) were the signals used in the analysis of the experiment.

5.3.5 Initial Signal Set

The initial signal set comprised of whistles that were produced using the restricted whistle. This meant that whistles were just as easily replicated in both conditions and were the same between conditions. The initial signal set was constructed from a set of signals which was designed to display a broad range of signal patterns and techniques. This was done in order to make the experiment comparable with the results of the original Verhoef experiment that also used an initial set that was designed to be unstructured. However, there are some problems with this approach, which I discuss at the end of the chapter.

There were 2 sets of initial signals. 2 chains in each condition started with the first set. 1 chain in both conditions started with the second set.

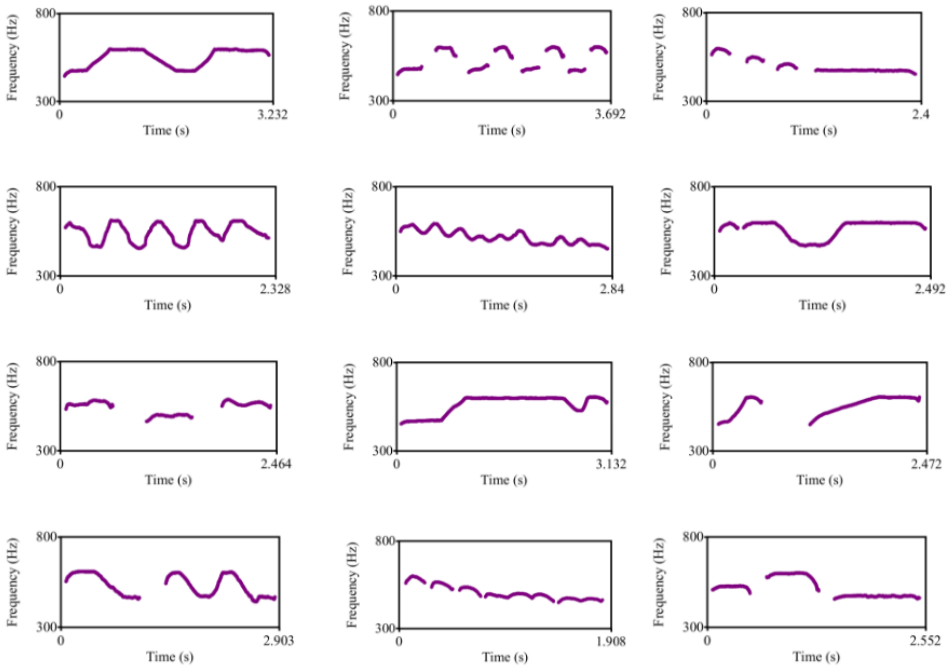

Figure 5.4: One of the initial signal inputs used in the experiment. Each signal is represented as a pitch track with frequency plotted against time.

5.3.6 Reproduction Constraint

As in the original Verhoef (2012) study, there was a constraint on signals that were produced in the recall phase. All signals had to be unique. This was done as in previous iterated learning experiments (i.e. Kirby et al., 2008) showed that without a pressure for expressivity (keeping signals distinct), then all signals end up the same. Indeed, the most learnable system is one where every signal is exactly the same. The pressure for expressivity in this experiment was an artificial pressure imposed by the software (designed and coded by Verhoef, 2012) that compared each signal after it had been produced with every signal already produced in that recall phase. If the signal was too similar to one already recorded, then the participant was asked to record the signal again. Similarity was measured using dynamic time warping (Sakoe & Chiba, 1978) and is detailed in Verhoef (2013).

5.3.7 Learning Bottleneck

There was no explicit bottleneck built in to the structure of this experiment. Participants heard all signals that they were to recall. However, the bottleneck came from the task being difficult. Nearly all participants expressed that they did not remember the whole signal set that they were asked to reproduce. The pressure to create new signals that fit in the language and were more "learnable", then, was the memory constraints of the participants.

5.4 Results

5.4.1 Features

For the analysis, I concentrated on the reuse of features and patterns as being the best indicator for the emergence of combinatorial structure. As a first step, I broke the whistles down into features. I did this manually. Signals were coded into strings of discrete features, listed in table 5.1. In most cases, the features were easy to isolate as they were separated by pauses in the whistle. However, this was not always the case. For example, if a whistle went up then down (a hill) and then up again (a rise) without a gap, this was not covered by the features in the table. In this example, this pattern could be coded as either a rise + a valley, or as a hill + a rise. In cases such as this, these patterns were always coded in the same way, in this example, as a hill + rise. Consistency in coding patterns allows for more reliable comparison of features and patterns between conditions and generations. Features that were coded by height of a pitch (peeps and plateaus) were done in respect to the pitch range available depending on condition, i.e. high peeps were in the top 33% of the signal space available depending on what that was, and the low peeps were in the the bottom 33%.

The first thing I did to measure the combinatorial structure in these feature strings was look at the occurrence rates of individual features across a repertoire of signals.

steep hill	long hill	steep valley	long valley
left skewed hill	right skewed hill	left skewed valley	right skewed valley
long rise	steep rise	long fall	steep fall
low plateau	high plateau	low peep	high peep
raising steps	falling steps	middle peep	middle plateau

Table 5.1: The features that were manually coded in the whistle signals. ”Steep” refers to quick changes in pitch, ”long” refers to slow changes in the pitch. Skewed hills and valleys are characterised by a one side of a hill or valley being steep and the other long. ”Plateaus” are static pitches. ”Peeps” are quick single notes. ”Steps” are 3 or more peeps rising or falling.

I calculated how many times each feature appeared in the signal repertoire of one participant (one generation in a chain). If a repertoire of signals has a feature or a small number of features that are repeated many times within the repertoire, then the mean of feature frequencies will be high. If there are many features only occurring a very small number times, then the mean will be low.

A mixed linear model, which included participant number and chain number as random effects, and generation and condition as fixed effects, found that there was no difference between the average feature frequency between the conditions. Comparing the mixed linear model with a null model found no significant effect ($\chi^2(1) = 0.34$, $p = 0.56$). However, looking at figure 5.5, in the restricted condition there seems to be a steady growth of average frequency of features. In order to investigate this trend, I split the data up between conditions. In the restricted condition, I did indeed find an effect of generation on the frequency of features ($\chi^2(1) = 7.04$, $p < 0.01$). In the unrestricted condition, there was no such trend ($\chi^2(1) = 0.06$, $p = 0.81$).

I also measured whether the features which were coded by the height of pitch (peeps and plateaus) were more frequent in the unrestricted condition as there was more room to make these distinctions. However, there was no relationship between the frequency of these features and condition ($t(31) = 0.6, p = 0.5$).

After this initial frequency analysis I looked more closely at patterns of features in the form of bigrams (groups of 2 features). For example, figure 5.6 shows the repetition of the bigram of 2 peeps which became more frequent as one chain in the restricted condition progressed. I calculated the frequency of bigrams across repertoires. 5.7 shows the average frequency of bigrams within each signal repertoire (one participant,

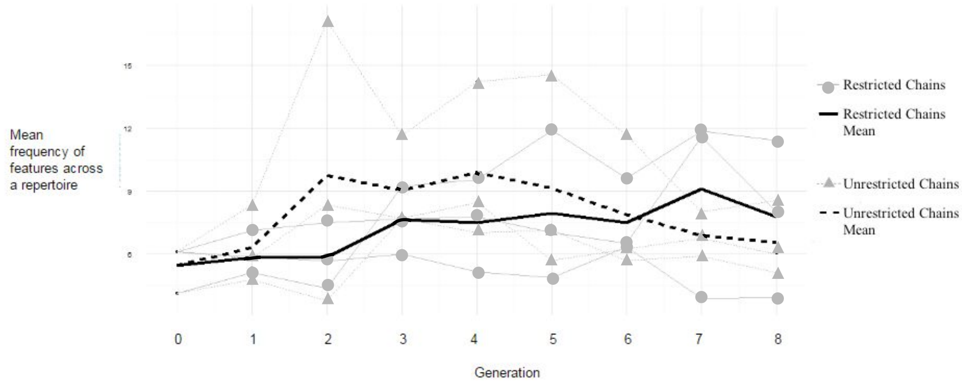

Figure 5.5: The mean number of feature frequencies within signal repertoires plotted against generation.

in one chain). The average was calculated using only the bigrams that appeared within a chain, not all possible bigrams. Again, using a mixed linear model, which included the same random and fixed effects as above, I found that there was no difference between the average number of bigrams repeated across a signal repertoire between the conditions ($\chi^2(1) = 0.13$, $p = 0.72$). But again, looking at the graph (figure 5.7), in the restricted condition, at least until generation 7, there seems to be a much more consistent growth of repeated bigrams across repertoires. I again split the dataset up between conditions. In the restricted condition, there is indeed an effect of generation on the number of bigrams ($\chi^2(1) = 6.82$, $p < 0.01$). However, in the unrestricted condition, there was no such trend ($\chi^2(1) = 0.14$, $p = 0.71$). Removing the outliers (see peaks in plot in figure 5.5) changed the results in the linear models slightly, but did not affect whether any of values reported above were significant or not.

5.4.1.1 Appearance of Novel Features

In the trends I have shown so far, there is not a consistent growth of repeated features or combined elements shown in the unrestricted condition. One reason for this may be that, because both conditions started with a set of signals produced using the restricted whistle, then with the unrestricted whistle there may have been an initial effect of innovation linked to participants wanting to explore the whole signal space. In an attempt to measure this, I calculated the number of whistle patterns that appeared in generations when they did not appear in the generation before. Again, with a mixed effect model with participant number and chain number as random effects, and generation and condition as fixed effects, there was no effect of condition on reinvention

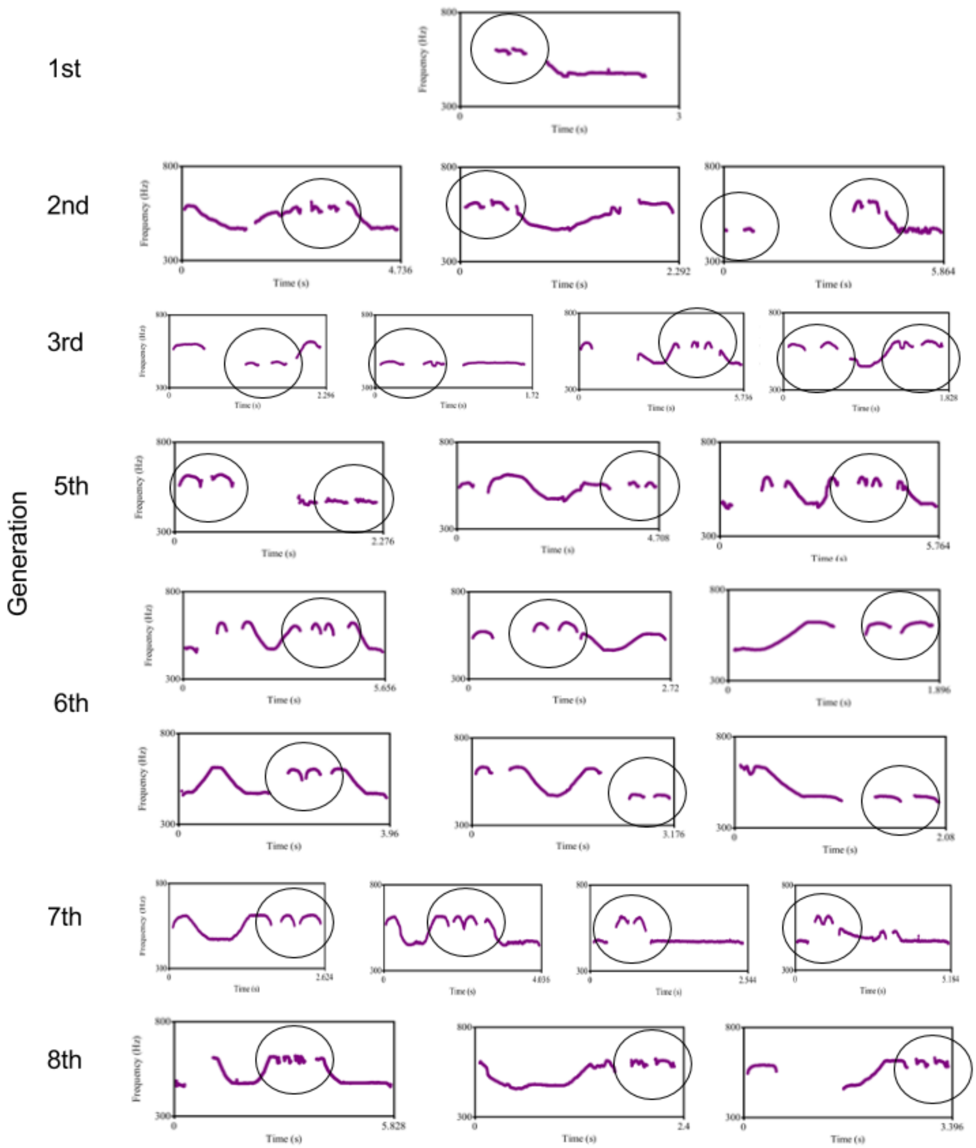

Figure 5.6: An example of a combinatorial pattern (2 peeps) growing in its frequency over generations in one chain in the restricted condition. Here both "high peeps" and "low peeps" are presented in the same figure.

rates ($\chi^2(1) = 0.64$, $p = 0.42$). Reinvention rates were also not affected by generation ($\chi^2(1) = 0.36$, $p = 0.55$).

5.4.2 Signal Duration

In chapter 2, I hypothesised that combinatorial structure will make use of the temporal dimension, with longer signals possibly facilitating the emergence of combinatorial

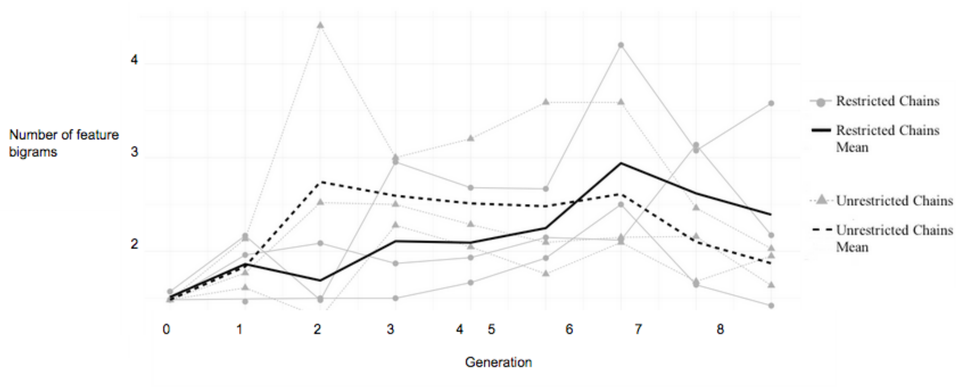

Figure 5.7: Number of element combinations repeated in participants' output within and across chains.

structure. If this is true, then the length of whistles may be a further indicator for the amount of combinatorial structure in a signal. Previously, Roberts et al. (2015) used duration as a measure for complexity, "because more complex signals are likely to take longer to produce" (p. 59).

I conducted a mixed linear model, which included participant number and chain number as random effects, and generation and condition as fixed effects, comparing with a reduced null model not containing condition. I found that length of whistle did not differ between the 2 conditions ($\chi^2(1) = 0.28$, $p = 0.6$). As can be seen in figure 5.8, there is a lot of overlap between the conditions, but both lines seem to be indicating that signals are getting longer in later generations. Comparing the same mixed linear model with a null model omitting generation, there was a borderline significant effect of generation ($\chi^2(1) = 3.44$, $p = 0.06$). This result hints at the use of temporal space in the whistles in later generations, possibly hinting that more combinatorial strategies are used in later generations. Having longer signals in the later generations may be directly related to the frequency of individual features being repeated within signals. Further, longer signals can have more features within them. Indeed, I found that including whistle length in the model for individual feature frequency above made the effect of generation disappear, indicating that whistle length and feature frequency are very dependent on each other. However, I do not think that this correlation negates either finding, as my hypothesis predicts that these features would correlate.

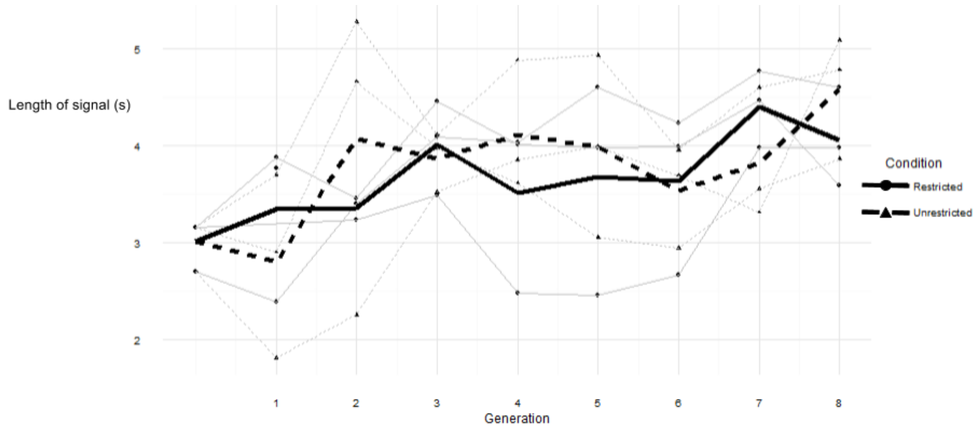

Figure 5.8: Average length of signals in seconds from participants' output within and across chains. The darker lines are the averages of chains in both conditions.

5.5 Discussion and Conclusion

I have presented an iterated learning experiment using slide whistles with two conditions: one a straightforward replication of Verhoef (2012), the other the same but with the whistle chamber, and therefor pitch range, restricted in its size. Using measures that were based on manual coding of patterns in the signals, I found no evidence that there was any difference between the conditions. However, there were some trends that existed in the restricted condition that did not exist in the unrestricted condition. Both frequencies of individual features and bigrams were predicted by generation in the restricted condition but were not in the unrestricted condition. Invention and reinvention of new features was not more likely in one condition than the other, neither was use of features that were more susceptible to signal space size differences. In both conditions, duration of signals was found to increase in later generations, but again, did not differ between conditions.

Taken together, the results show some interesting trends that indicate that there may possibly be differences caused by the smaller signal space. However, due to no significant differences found between the conditions in any of the dependent variables, I cannot conclusively claim that signal space size had an effect in this experiment. There are several possible reasons for this that stem from the experimental design. I detail these below. I have already covered in Chapter 4 some problems with using slide whistles, and so I will save repetition here.

5.5.1 Replication of Patterns

The aim of the manipulation in this experiment (signal space size) was to limit the size of the signal space. However, how units are defined within this paradigm (and the original Verhoef study) is not necessarily in terms of pitch, but rather in terms of pitch patterns and stopper movements. If I was only measuring signals in terms of pitch, I would find that there is a smaller pitch range used in the restricted condition because that was part of the experiment design. Patterns are not necessarily affected by the restriction of the chamber size in the whistle. It is arguably just as easy to create a rising or falling pitch with a restricted whistle as it is with the unrestricted one. As a result, using pitch trajectories as analogous to phonemes, as Verhoef et al. (2014) does, it is less likely that the possible inventories will be affected by the size of the whistle chamber. However, other elements of the signal space may affect crowding in the signal space, such as dimensionality, which is explored in the next chapter.

5.5.2 Optimally Unstructured Initial Input

Other problems lie not so much in the design of this particular experiment, but the original Verhoef (2012) study. In the Verhoef study, all participants in the first generation of the chains received the same initial set of signals. These signals were produced from a set of signals by naïve participants who freely created whistles from scratch. The set of signals that were learnt in the first generation were selected from this set. They were selected to have a wide range of variability. This was done so that "the total set of whistles did not exhibit any observable combinatorial structure" (p. 4). This causes some problems with interpreting the results, especially in early generations. Some whistle patterns or techniques are likely to be more prevalent than others when people are producing whistles randomly. Frequencies will be dependent on ease of articulation or pre-existing biases about music or language. The fact that whistles were chosen to be in a set without combinatorial structure suggests that if signals had been chosen from the set of freely produced signals at random then the resulting set would have had some combinatorial structure (with some signal elements repeating across the set). This potentially sets the baseline of combinatorial structure found in the initial whistle sets to be lower than one might expect from chance, meaning that we should perhaps not be too surprised by the emergence of combinatorial patterns until it reaches past the point that it would occur from chance. This chance level is highly dependent on articulation constraints. As I pointed out in Chapter 3, some trajectories are more likely than others, for example, once a participant has reached the lowest note with the plunger all the way out. Ease of articulation is also likely to influence the combinatorial structure that emerges past this chance level.

5.5.2.1 Oversensitive Reproduction Constraint Algorithm

Participants often reported that the algorithm that prevented reproductions of already produced signals was somewhat overzealous, rejecting signals which they were certain they had not already produced. It's difficult to know quite what was causing this, as rejected whistles were not recorded by the software, and participants only reported problems once the experiment was over. This over-zealousness was not corrected during the experiment, as the same code was used for the original Verhoef (2012) experiment, and it was important that this study be a replication in order to be comparable. It is perhaps the case, then, that instead of promoting signals that "fit the language well", the algorithm might have provoked signals that were more distinct so that participants could be sure that it would not reject further signals. This could have possibly acted against producing results that would have had a high degree of combinatorial structure.

5.5.3 The Lack of Meanings

The lack of meanings in this experiment was meant to promote bare phonology rather than combinatorial structure, as it is defined in relation to meaning (see Chapter 2 for discussion of this). Investigation of bare phonology was used as justification for not having meanings in Verhoef (2012). In Verhoef et al. (2012), the authors state (of the slide whistle experiment) that "the goal was to simulate the emergence of 'bare phonology'" (p.370), as well as preventing structure in the semantic space being a confound. However, this experimental design choice relies on the assumption of the existence of bare phonology, which perhaps assumes that the emergence of phonology predates compositional structure. This means that phonological structure would need to emerge before the need to attach the resulting signals to meanings. Such a theory would rely on something like musical protolanguage existing before language was co-opted for communication. While this theory has its merits, it is not ubiquitous in the language evolution literature and so remains a major assumption. Further, it is difficult to justify labelling structure as combinatorial without relation to meaning. Defining structure as meaningless does not provide insight if there is no meaning. Accordingly, all of the experiments in the rest of this thesis will be using meanings.

One defence for the strength of the original Verhoef (2012) experiment lies in the fact that structure emerges even in the absence of meaning or pressures for communication (though the reproduction constraint created an artificial proxy for expressivity). This shows that vertical transmission alone is possibly sufficient to have structure emerge.

5.5.4 Quantitative analysis

The analysis I used in this chapter was highly qualitative in nature. I manually coded the signals myself, and as the experimenter I will have had biases when it came to the decisions I made. A system that finds structure automatically without too many priors is much more preferable. In her experiments, Verhoef used associative chunk strength (similar to the bigram counts here) and entropy. Entropy was quite difficult for me to measure here, as there was variability in the quality of recordings (sound proof booth verses classroom) which meant comparison was difficult. This was one of the reasons that the Leap Motion paradigm Eryılmaz & Little (2016) is cleaner as it records the data that can reconstruct signals, not the audio produced. Accordingly, the experiments in the rest of the thesis, which use Leap Motion, are much more quantitative in their analysis.

CHAPTER 6

Signal space dimensionality[1]

In this chapter, I present Experiments 2 and 3, which are both artificial signal creation experiments using the Leap Motion. They test the effect of signal space dimensionality on the emergence of combinatorial structure. Experiment 2 compares signal spaces where either pitch or volume or both refer to meanings that either differed in only size, or both size and shade of grey. However, I found that participants used duration as a signal dimension, meaning that the number of signal dimensions did not correspond to the intended number in the experimental design. To fix this, in Experiment 3, signals only differ in pitch (and duration) and the meaning space grows to 3 dimensions to ensure we can observe the effects of meaning dimensions outnumbering signal space dimensions.

6.1 Introduction

Experiment 1 in Chapter 5 investigated how the size of signal space affected the emergence of combinatorial structure. However, one issue with Experiment 1 was that the physical size of the space did not reduce the number of potential distinctions which could be made using the signal space. In this chapter, I focus on the dimensionality of a signal space which is defined as the number of ways distinctions can be made.

I have already discussed dimensionality in Chapter 2, outlining the theoretical reasons why signal space dimensionality should affect the emergence of combinatorial

[1]The work in this chapter is published in Cognition (Little et al., 2017b).

structure, and in Chapter 3, where I discussed how signal space dimensionality may relate to modality. In the experiments presented in this chapter, dimensionality is experimentally manipulated using the Leap Motion. Before outlining the experiments and their results, I will outline how the issues raised in Chapters 2 and 3 relate to the signal space being used in these experiments.

In chapter 2, there were 2 hypotheses for why signal space dimensionality might affect the emergence of combinatorial structure. The first was that crowding on individual dimensions creates a pressure for signals to be more distinguishable from one another, possibly through using combinatorial structure. The second was that the number of meaning dimensions outnumbering the signal dimensions will cause a) the signal space to become crowded, and b) reduce the ability to produce iconic signals. The experiments presented here have their primary focus on the second hypothesis. The experiments are not designed to tease apart points a and b, as these effects are possibly not separable. The first hypothesis is touched upon briefly in the discussion, but further experimental work is needed to explore crowding on individual dimensions.

The experiments here are designed to directly test the results of the computational model by de Boer & Verhoef (2012). This model demonstrated that when the number of signal dimensions matches the number of meaning dimensions then signals emerge that are high in iconicity and low in structure. The model also shows what happens when the meaning dimensions outnumber the signal dimensions, which is that signals develop structure. The dimensions in my experiments are manipulated by making the space above the Leap Motion either 1 dimensional or 2 dimensional. Here, dimensionality is being interpreted in a very literal way from mathematical dimensionality as it is in the de Boer & Verhoef (2012) model. However, I encountered some problems with this method with human participants, which I resolve in Experiment 2. Discussion on what the signal space looked like exactly are detailed in the methods below.

In these experiments, both the signal dimensions and the meaning dimensions are continuous and are designed to facilitate diagrammatic iconicity. For instance, one signal dimension is pitch and one meaning dimension is size. These can be mapped onto each other relatively easily. For instance, participants may create a system where higher pitches refer to smaller referents. The possibility for diagramatic iconicity sets this experiment apart from previous experiments that have explored how iconicity affects the emergence of combinatorial structure. Previous attempts (Roberts & Galantucci, 2012; Roberts et al., 2015; Verhoef et al., 2015a) operationalise iconicity as something more similar to imagic iconicity where it is not necessary to see a full system of signals and meanings in order to know the iconicity is present. Not all iconicity in language is imagic, and so my experiment brings a new perspective on how iconicity can influence the emergence of structure.

6.2 Experiments

My hypothesis for both Experiments 2 and 3 is that when the dimensionality of the signal space is lower than that of the meaning space, then a more structural strategy will be adopted. I also hypothesise that when there is a transparent mapping possible between the dimensionality of the signal space and the meaning space (because their dimensionalities match), then iconic strategies will be adopted.

6.3 Experiment 2

Experiment 2 is a signal creation experiment of the sort outlined in chapter 4, comprising of signal creation tasks and a signal recognition tasks. Participants created signals that they recognised themselves. These signals were not subject to communication between participants, or iterated learning.

6.3.1 Methods

6.3.1.1 Participants

Participants were recruited at the Vrije Universiteit Brussel (VUB) in Belgium. 25 participants took part in the experiment; 10 male and 15 female. Participants had an average age of 24 (*SD* = 4.6). I also asked participants to list their native language, as well as other languages with their level of fluency. No participants reported any knowledge of sign languages. I also asked participants to self-report their musical proficiency (on a scale of 1-5). This information was recorded as recognition of pitch-track signals might be dependent on participants' musical abilities, so I needed to identify and control for this potential effect in our results.

6.3.1.2 The signal space

The experiment used the Leap Motion sensor, as detailed in Chapter 4. Participants could either manipulate signals by moving their hand within a horizontal dimension

(x) to affect pitch or the vertical dimension (y) to affect volume or both (Figure 4.1). Both the pitch and volume scales used were non-linear (again outlines in Chapter 4).

6.3.1.3 The meaning space

The meaning space consisted of a set of squares that differed along continuous dimensions. Participants had to create distinct signals for each square. In phases where the meaning space only differed on one dimension, five black squares only differed in size. In phases where the meaning space differed on two dimensions, nine squares differed in both size and in different shades of grey (Figure 6.1).

6.3.1.4 Procedure

Participants were given instructions on how to generate signals using the Leap Motion. They were given time to practice using the Leap Motion as long as the instruction screen was showing. The participants had control of when they decided to start the experiment. They were instructed to sit back in the chair during the experiment, so that their upper body did not interfere with the Leap Motion. Participants were also told that they would have to recognise the signals they produced, so they knew they had to make signals distinct from one another.

There were three phases of the experiment, each phase consisted of a practice round and an experimental round. There was no difference between practice rounds and experimental rounds, but only the data from the experimental round was used in the analysis. Each practice and experimental round consisted of a signal creation task and a signal recognition task.

6.3.1.5 Signal Creation Task

At the beginning of each signal creation task, participants saw the entire meaning space for that phase. They were presented with squares in a random order, one by one, and clicked an on-screen button to begin and finish recording their signals. They had the opportunity to play back the signal they had just created, and rerecord the signal if they were not happy. Participants were told explicitly which signal dimension(s) they were manipulating. Participants created signals for all possible squares in a phase.

6.3.1.6 Signal Recognition task

After each signal creation task, participants completed a signal recognition task for every signal they had created in a random order. They heard each signal, and were asked to identify its referent from an array of four randomly selected possibilities (from the repertoire of possible meanings within the current phase). They were given immediate feedback about whether they were correct, and if not, what the correct meaning had been. This task worked as a proxy for the pressure of expressivity, as participants knew that they had to produce signals that they could then connect back to the meaning in this task, thus preventing them from producing random signals, or just the same signal over and over again. Their performance in this task was recorded for use in the analysis. When participants were incorrect, I also measured the distance between the answer they gave and the correct answer. The distance was calculated as the sum of differences along each dimension using a measure similar to Hamming distance defined below.

Let m_{ij} define a meaning with size i and shade j in a meaning space where $0 < i < I$ and $0 < j < J$. The distance between two meanings m_{ij} and $m_{i'j'}$ is then the following:

$$D(m_{ij}, m_{i'j'}) = |i - i'| + |j - j'| \tag{6.1}$$

So, for example, if the correct square has values 3 and 3 for size and shade respectively, and the chosen square had vales 1 and 2 for size and shade respectively, the distance between these two squares would be 3. Correct answers will have a distance of 0.

6.3.1.7 Phase 1:1

All participants started with phase 1:1. In this phase, the meaning space consisted of five black squares, each of different sizes (one meaning dimension). In this phase, the signal space had also only one signal dimension, which was either pitch or volume. Which signal dimension the participants started with was assigned at random. This phase was a matching phase, as there was a one-to-one mapping possible between the meaning space and signal space (Figure 6.1).

6.3.1.8 Phases 1:2

In phase 1:2, participants created signals for a two-dimensional meaning space with the squares differing in size and shade. The signal space had only one dimension. Participants used the same one-dimensional signal space which they used in phase 1:1, so if they started the experiment only using pitch, they only used pitch in this phase. This was the mismatch phase, as there were more meaning dimensions than signal dimensions (Figure 6.1).

6.3.1.9 Phases 2:2

In phase 2:2, participants described the two-dimensional meaning space (differing in size and shade), but with a two-dimensional signalling space, where the signals differed in both pitch and volume along the x and y dimensions respectively (Fig. 6.1). This phase was a matching phase also, as there was a one-to-one mapping available between signal and meaning spaces.

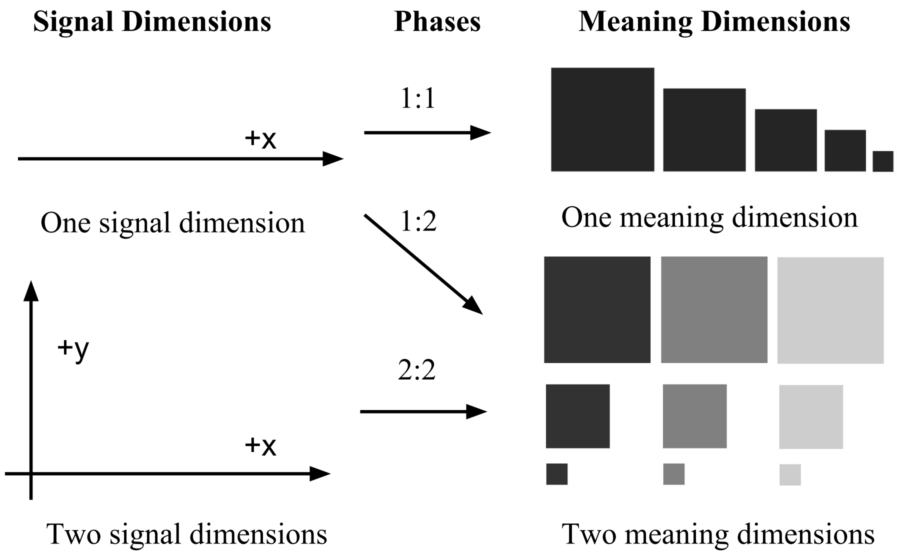

Figure 6.1: The phases used in Experiment 2.

6.3.1.10 Counterbalancing

Participants completed the phases in order 1:1, 1:2, 2:2 (so mismatch phase interrupts matching phases) or 1:1, 2:2, 1:2 (where matching phases are consecutive). The order was randomly assigned. Participants' behaviour may depend on what they have previously done in the experiment. If people have to solve the dimensionality mismatch before being presented with the two-dimensional signal space, then they may continue to express meanings in the two-dimensional signal space with an already established strategy using one dimension, rather than change their strategy to take advantage of both dimensions.

6.3.1.11 Post-experimental questionnaire

I conducted a questionnaire with each participant after they had completed the experiment. This questionnaire asked about the ease of the experiment, as well as about the strategies that the participant adopted during each phase of the experiment. The questionnaire was free-form, but asked explicitly whether they had a strategy and, if so, how the participant encoded each meaning dimension into their signal. The questionnaires were primarily used in order to know whether from the participant's point of view signals were iconic and whether, if they had structure, the structure was characterised as being combinatorial or compositional from the perspective of the participant.

6.4 Results

6.4.1 Signal Creation Task

The data collected from the signal creation task consisted of coordinate values designating hand position at every time frame recorded, which is what the following statistics are based on. Signals were on average 370 frames long (approx. 3.36 seconds[2]). Due to the complexities involved in generating quantitative measures for structure in continuous signals, I first looked at simple descriptive statistics, such as the mean of the coordinate values for each trajectory, and the duration of a signal. These simple measures gave me a good starting point to assess whether participants were encoding the meaning space directly with the signal space. If size or shade was directly encoded by pitch, volume or duration in the form of diagrammatic iconicity, then this should be detectable in the mean coordinates or duration of the trajectories.

The first dimension seen (pitch or volume) was treated the same within the analysis. All coordinates for signals using either pitch or volume were transformed to have the same range, though I also controlled for whether these coordinates were pitch or volume in the mixed linear models below and also ran a separate analysis that showed that participants performed just as well in the task when starting with either pitch or volume (reported in the signal recognition results below). As explained above, meaning dimensions were coded to reflect the continuous way in which they differed, i.e. the smallest square was coded as having the value of 1 for size, and the biggest square

[2]The frame rate of the Leap sensor is variable, from about 30 frames per second (fps) up to about 120 fps. However, it is heavily centred around the median 110 fps, with over 75% of the frames captured at above 90 fps. This variability was fixed in later experiments presented in this thesis.

a value of 5, with the lightest grey square given a value of 1 for shade, and the darkest had a value of 3. Using these values, I could correlate size and shade with duration and mean coordinates.

In the first phase, only duration reliably predicted the size of the squares ($F(1,103) = 6.8, p = 0.01$). In the other 2 phases, the mean coordinate of signals on the first dimension that a participant used in phase 1:1 (either pitch or volume) was predicted most strongly by shade. A mixed linear model, which included participant number as a random effect, and whether their starting dimension was pitch or volume as a fixed effect, showed this result to be significant ($\chi^2(1) = 341.4$, $p < 0.001$). The duration of the signal was predicted most strongly by the size of the square, with each step of size increasing the signal by 75.296 frames$\pm$7(std errors) (approx 0.7 seconds, see section 4.5.1 for frame to second conversion). The mixed linear model for this interaction, controlling for the same fixed and random effects, was also significant ($\chi^2(1) = 103.14$, $p < 0.001$). These correlations demonstrate a propensity for encoding the meaning space with the signal space using diagrammatic iconicity. Size and duration are easy to map on to one another, and it makes sense that participants will likely encode the remaining meaning dimension (shade) with the signal dimension they were first exposed to. Figure 6.2 shows the output of one participant who mapped the signal space on to the meaning space in a very straightforward one to one mapping, with size encoded with duration and shade encoded with volume.

Further to the basic statistics above, standard deviation scores gave me a good idea of the amount of movement in a signal. Signal trajectories produced in the phase where there was a mismatch (1:2) had higher standard deviations (48.2mm) than signals produced in phases where the signal and meaning spaces matched in dimensionality (33.4mm) , indicating more movement, and therefore structure, in mismatch phases. Using a linear mixed effects analysis and controlling for participant number as a random effect, and whether they started with pitch or volume as a fixed effect, this finding was significant ($\chi^2(1) = 4.5$, $p < 0.05$)

6.4.2 Predictability of signal trajectories

I measured the predictability values for each trajectory within a repertoire (summary of method is in Section 4.5.3), I created a mixed effects linear model comparing the predictability of signals in a matching and mismatching phases. Controlling for duration as a random effect, and size of square and participant number as fixed effects, I found

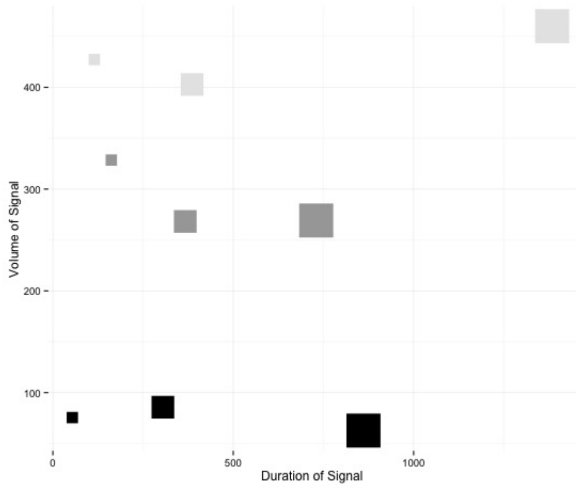

Figure 6.2: The mean trajectory coordinates (in mm) along the axis manipulating volume (where lower values refer to louder sounds) plotted against duration (number of data frames, roughly 1/110 of a second). Size and shade are represented by the size and shade of the squares in the graph. Within the phase with the two-dimensional meaning space with a two-dimensional signal space, this participant used signal duration to encode size, and signal volume to encode shade.

that whether signals were produced in matched or mismatched phases predicted how predictable a trajectory was ($\chi^2(1) = 3.9$, $p < 0.05$). Signals produced in the matching phases had higher predictability.

6.4.3 Hidden Markov Models

As mentioned briefly in chapter 4, a fellow PhD student, Kerem Eryılmaz used Hidden Markov Models (HMMs) to characterise structure in signal repertoires produced in this experiment by calculating the number of hidden “states” in signals across a repertoire. The idea behind these models is that the states are building blocks that make up the signals. These models were built to measure combinatorial structure. If a repertoire of signals relies on diagrammatic iconicity, then you would expect a lot of states to model such a system, because every meaning along a meaning dimension will need its own state. However, if signals are produced using more combinatorial strategies, then fewer states are expected because signals will be composed from a small, finite set of reusable building blocks. Using this data, we can test if the number of states predicted participants’ ability to recognise their own signals, and to see if repertoires were more likely to have fewer states when there was a dimensionality mismatch.

First, recognition success was most reliably predicted by the number of states when the models were trained using the transformed Mel frequency values for pitch and amplitude in linear scale for volume. Accordingly, the results in this section are derived from HMMs trained using those projections.

Results from the HMMs found that if phases were presented with 1:2 before 2:2 participants often changed strategy to take advantage of both dimensions and increased the number of states, even though they could in principle keep the same strategy in 2:2 as they used in 1:2. Changing from 2:2 to 1:2 also resulted in an increase in the number of states required. When the 2:2 phase was presented before 1:2, it required fewer states than in any other phase in any other order, even the 1:1 phase. These trends, however, were not significant. For both orderings, changes in the number of states between phases 1:2 and 2:2 was insignificant, suggesting no evidence for more combinatorial structure when phases were mismatching phases. However, the number of states overall was fewer when phase 2:2 appeared before phase 1:2. More details can be found in Kerem Eryılmaz's PhD thesis.

6.4.4 Signal Recognition Task

There was no significant difference between the recognition rates between participants who were assigned to start with either volume or pitch (t(21.9) = -0.46, $p = 0.65$), suggesting that there was no difference between the difficulty of either signal dimension.

Overall, participants were good at recognising their own signals, identifying a mean of 66% of signals correctly. 25% was expected if participants performed at chance level. Using a linear regression model, I found that participants improved in their performance throughout the experiment, with the performance at recognising their signals improving by around 10% with each phase (F(1, 76) = 9.96, $p < 0.01$).

If signals rely on diagrammatic iconicity, then similar signals used for similar squares will be more easily confused than signals for less similar squares. This confusability in systems with diagrammatic iconicity may cause participants to be worse at the signal recognition task when diagrammatic iconicity is more prevalent. In line with this hypothesis, I found that participants were worse at recognising their signals within matching phases (1:1, 2:2) (61.3% correct, s.d. 24%), than in mismatching phases (1:2) (69.6%, s.d. 21%). However, this result was not significant (t(53.3) = -1.5, $p = 0.13$), and may be an artefact of the experiment getting easier as the experiment progressed, as omitting the data from the first phase made the difference between matching and mismatching phases only 2%.

I also calculated the distances between incorrect answers and target answers, as discussed in the methods section above. In order to put these values in context, I also calculated the distance between the target answer and an incorrect answer which was chosen at random from the array of 3 possibilities from the real experiment (using the random module in Python). The correct answer could not be chosen in order to allow for comparison with the actual data, which only included incorrect answers for this analysis. Comparing the actual data with the random data using a mixed effects linear model, and controlling for participant number as a random effect, and stimulus number as a fixed effect, I found that with incorrect choices produced in the matching phases (1:1, 2:2), participants were closer to the correct square (2.6 steps away, s.d. 1.4) than if they had chosen at random (3 steps away, s.d. 1.7) ($\chi^2(1) = 5.5$, $p = 0.02$). However, in the mismatching phase (1:2) there was no difference between actual incorrect choices and random incorrect choices (both around 3.6 steps away, $\chi^2(1) = 0.01$, $p = 0.9$). However, there was not a significant interaction between whether an observation was matching or randomly generated ($\chi^2(1) = 1.3$, $p = 0.25$). Further, I found that the distance from the correct answer was much higher in the mismatching phases (3.6 steps away, SD = 1.5), than in the matching phases (2.6 steps away, SD = 1.4), indicating that participants were relying more on diagrammatic iconicity in the matching phases, because their mistakes were predicable from assuming a transparent mapping between the signal space and the meaning space. I tested this using a mixed effects linear model, and controlling for the same variables ($\chi^2(1) = 5.3$, $p < 0.05$).

I used a linear regression model to test if musical proficiency predicted performance in the signal recognition task, and found that it did not ($F(1, 23) = 0.03$, $p = 0.86$).

6.4.5 Post-experimental questionnaire

Despite the questions being free-form, it was easy to extract meaningful data from the post-experimental questionnaires. Across all participants, there was a relatively small number of strategies self-reported. These strategies included using pitch, volume or duration directly to encode size or shade. For example, using high pitches or short durations for small squares and low pitches or long durations for big squares. There were also strategies reported which involved different movement types, frequencies and speeds.

As I predicted in the section on counterbalancing, participants who saw phase 1:2 before phase 2:2, were more likely to report using the same signalling strategy throughout, than to change the strategy to take advantage of both dimensions (though see HMM section that this might not have been 100% the case). 84% (s.d. 37%) of strategies used for a particular meaning dimension were consistent throughout the experiment by participants who saw 1:2 first. Only 54% (s.d. 50%) of strategies by those who saw 2:2 first were consistent. This association was significant ($\chi^2(1) = 8.7$, $p < 0.01$).

Whether a participant self-reported as having a strategy or not also influenced their performance in the signal recognition task. Participants were significantly more likely to perform better at recognising their own signals in a given phase, if they had a strategy (M=70% correct, s.d. = 20%), than if they didn't have a strategy (M=40% correct, s.d. = 16%) (t(26.6) = -6, $p > 0.001$).

6.5 Experiment 3

Experiment 2 provided some important evidence pertaining to what happens when there are different mappings between signal and meaning spaces. When there is a one-to-one mapping available between signal and meaning spaces, participants tend to take advantage of it. Indeed, even in the conditions designed to produce a dimensionality mismatch, participants used duration as another signal dimension. Despite this, I was still able to find significant effects of the matching phases compared to the mismatching phases.

Experiment 3 is a very similar signal creation experiment. It tested the same hypothesis as Experiment 2, but the design was altered to counter two possible problems with Experiment 2:

1) That duration was used as a dimension by some participants, meaning that if participants used duration, there wasn't really a "mismatch" even with the 1:2 phase.

2) Participants created signals for a very small and finite meaning set in Experiment 2 (5 or 9 meanings depending on the phase), which they saw in its entirety before the experiment. This facilitated the ability for participants to create a completely holistic signal set, without the need for structure in order for the signals to be generalisable from other signals in the set. Only one participant treated meanings holistically in Experiment 2 (using frequencies of pitch contours to differentiate meanings). However, I feel that this is still a flaw in the experimental design, as this strategy would

soon become maladaptive as meaning numbers rise, and so is inconsistent with what would happen in analogous situations in the real world, where continuous meaning dimensions are much more nuanced than having 3 or 5 differences.

In order to combat these problems, two alterations have been made:

1) Phase 1:2 in Experiment 2 has been dubbed a "mismatch" phase, and a new phase 1:3 has been instated to be sure there is a dimensionality mismatch.

2) There is now a meaning bottleneck, where participants do not create signals for every possible meaning, but a subset of them, this is explained further in the *Meanings* section below.

6.5.1 Methods

6.5.1.1 Participants

Participants were recruited at the VUB in Brussels. 25 participants took part in the experiment; 8 male and 17 female. Participants had an average age of 21 (SD = 3.2). As in Experiment 2, I asked participants to list the languages they speak, with level of fluency, and to self report their musical proficiency (on a scale of 1-5).

6.5.1.2 Signals

As in Experiment 2, there was a continuous signal space which used the *Leap Motion* sensor to convert hand motion into sounds. However, in this experiment, signals could only be manipulated in pitch. Participants manipulated the pitch in the same way as in Experiment 2, along the horizontal axis, and with a logarithmic relationship between hand position co-ordinates and signal frequency. The vertical axis was not used at all in this experiment, meaning that, including duration, the number of signal "dimensions" could not be more than 2. However, participants were not explicitly told to use duration. Again, participants were given clear instructions on how to use the sensor, and were given a practice period to get used to the mapping between the position of their hand and the audio feedback before the experiment started.

6.5.1.3 Meanings

The meaning space again consisted of a set of squares that differed along three continuous dimensions: size, shade of orange, and shade of grey. Participants had to create distinct signals for each square they saw. Squares differed along different numbers of dimensions in each phase (Figure 6.3). In contrast to Experiment 2, the number of possible squares from the meaning space outnumbers the number of squares presented to the participants. Each dimension differed by 6 degrees, meaning that the signal space grew exponentially with the number of dimensions (see description of phases below). Having 6 degrees of difference on meaning-space dimensions meant the meaning space is big enough in order to have an effective bottle-neck, but small enough to not make the discrimination task impossible. Further to the reasons given above, this aspect of the experimental design created an incentive for participants to create more productive systems that extend to meanings they have not seen yet. This pressure caused by a meaning bottleneck has been shown to encourage the production of structure in experiments such as Kirby et al. (2008) and Kirby et al. (2015).

Two of the meaning dimensions in this experiment, were "shade of grey" and "shade of orange". In pilots, I originally had the squares differ in shade of orange (which I affected using the RGB ratio of green to red) and the brightness value. However, this made the squares at the darker and redder end of the scale very difficult for participants to tell apart, as they all appeared the same dark brown colour. To solve this, I used striped squares with alternating grey and orange stripes (see figure 6.3). This gives the same effect of squares differing in shade of orange and brightness, but squares at both ends of the spectrum can be distinguished just as easily.

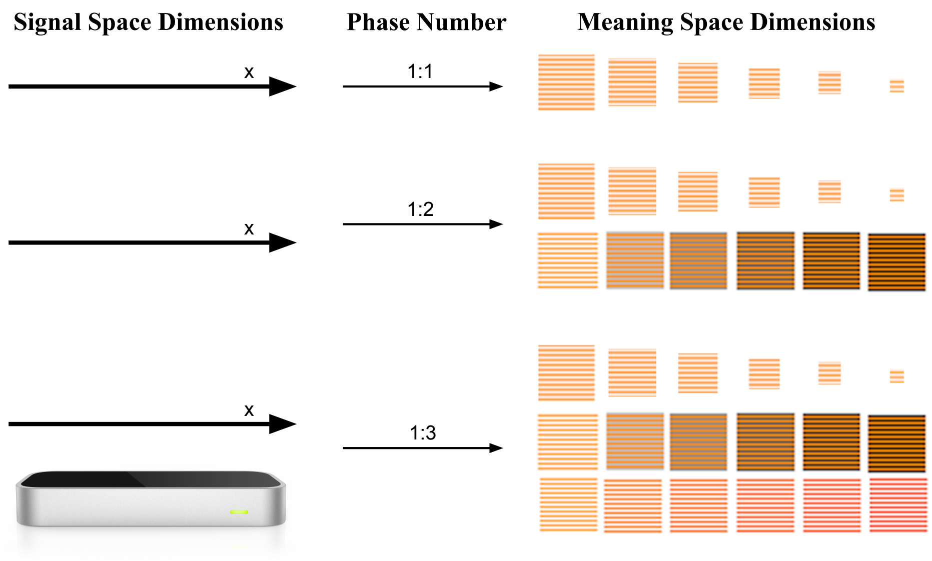

Figure 6.3: The dimensions used in Experiment 3 in each of the 3 phases.

6.5.1.4 Procedure

The procedure in Experiment 3 was nearly the same as Experiment 2. There were still 3 phases, each with a practice round and an experimental round, which were both the same. Each round has a signal creation task and a signal recognition task. However, the phases were slightly different.

6.5.1.5 Phases

All participants had phases presented in the same order: 1:1, 1:2, 1:3. The “1” here refers to 1 signal dimension (pitch), in order to make these phase labels consistent with the phases in Experiment 2. However, since I have learnt to expect participants to use duration as a signal dimension, it is important to remember that the meaning dimensions only outnumber the signal dimensions in a meaningful way in phase 1:3.

6.5.1.6 Phase 1:1

In phase 1:1, there were 6 squares which differed in 6 degrees of size. All 6 squares were presented in a random order.

6.5.1.7 Phase 1:2

In phase 1:2, participants were presented with 12 squares in a random order, which differed along two dimensions: 6 degrees of size and 6 shades of grey stripes (See Figure 6.3.) This made a possible number of 36 squares that were chosen from at random.

6.5.1.8 Phase 1:3

In phase 1:3, participants were presented with 12 squares in a random order, which differed along three dimensions: 6 degrees of size, 6 shades of grey stripes and 6 shades of orange stripes (See Figure 6.3.) This made a possible number of 216 squares that were chosen from at random.

6.5.1.9 Signal Recognition task

As in Experiment 2, participants completed a signal recognition task. They heard a signal they had created, and were asked to identify its referent from an array of four randomly selected squares from the set of possible squares in the current phase. They were given immediate feedback about whether they were correct, and if not, what the correct square had been. Their performance in this task was recorded for use in the analysis. How far away in the meaning space they were from the correct answer was also recorded in the same way that it was in Experiment 2.

6.5.1.10 Post-experimental questionnaire

The experiment also had a post-experimental questionnaire. This questionnaire asked about the strategies that the participant adopted during each phase of the experiment. As in Experiment 2, the questionnaire was free-form.

After the experiment, participants were also given a colour chart of the shades of orange (Figure 6.4). The shades used in the experiment had been designed to all be perceived as orange, though linguistic knowledge will obviously affect this categorisation. This part of the post-experimental questionnaire was designed to see if participant's categorisation of these colours affected their signals and performance. Only 17 participants completed this later part of the questionnaire because of experimenter error.

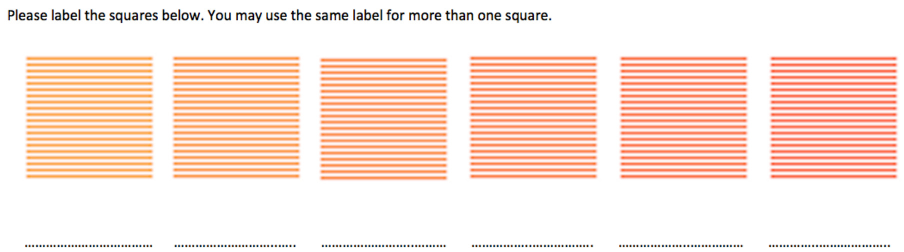

Figure 6.4: The categorisation task participants were asked to complete.

6.6 Results

6.6.1 Descriptive Statistics

In this experiment, signals were on average 252 frames long (approx 2.3 seconds). The average duration of signals rose by about 20 frames each phase ($\chi^2(1) = 7.9$, $p < 0.005$). As in Experiment 2, meaning dimensions were coded to reflect the continuous way in which they differed, i.e. the smallest square was coded as having the value of 1 for size, and the biggest square a value of 6, with the lightest grey stripes were given a value of 1 for colour, and the darkest had a value of 6. Again, across all phases, the size of square was the best predictor for the duration of the signal ($\chi^2(1) = 63.3$, $p < 0.001$), with the smallest squares having a mean duration of 1.55 seconds (SD = 1.26s), and the largest squares having a mean duration of 2.7 seconds (SD = 1.9s). However, in this experiment, size was also the best predictor for the mean pitch of the signals ($\chi^2(1) = 15.7$, $p < 0.001$). With the smallest squares having a mean pitch of 403Hz, and the largest squares having a mean pitch of 333HZ. Again, I take this as evidence for the use of diagrammatic iconicity.

I again looked at the standard deviations of individual signal trajectories to see if the degree of mismatch in the signals affected the amount of movement in the signals. There was no significant difference between phases where there was no mismatch (Phases 1:1 and 1:2), in fact, the mean standard deviation in these phases was nearly identical (around 28mm, SD = 31.5). However, the SDs from phase 1:3, where there was definitely a mismatch, was significantly higher (M = 33.8mm, SD = 34.4) than in the other two phases ($\chi^2(1) = 6.9$, $p < 0.01$). Figure 6.5 shows how this effect manifested itself in the signals of one participant where the differences between phases was particularly marked.

6.6.2 Predictability of signal trajectories

I again calculated the predictability values for each of the signal trajectories in a repertoire in the same way as we did in Experiment 2. I was interested to see if the phase (or the amount of signal-meaning mismatch there was) had an effect on how predictable the signals were. Using a mixed effects linear model and controlling for duration and participant number as a random effect, and size of square as a fixed effect, I found that whether the signal was produced in a matching phase or not correlated with how predictable a trajectory was ($\chi^2(1) = 11.2$, $p < 0.001$). The value was closer to 0 (so more predictable) in phase 1:1 (mean = 95), and got less predictable with each phase (phase 1:2 mean = 119, phase 1:3 mean = 145).

6.6.3 Hidden Markov Models

HMMs did not produce any significant results for Experiment 3. Why they did not work is not at all clear and all I can do is speculate about the most salient differences between Experiments 2 and 3. For instance, the two-dimensional signal space in Experiment 2 may have worked particularly well with the HMMs, or perhaps the 1:3

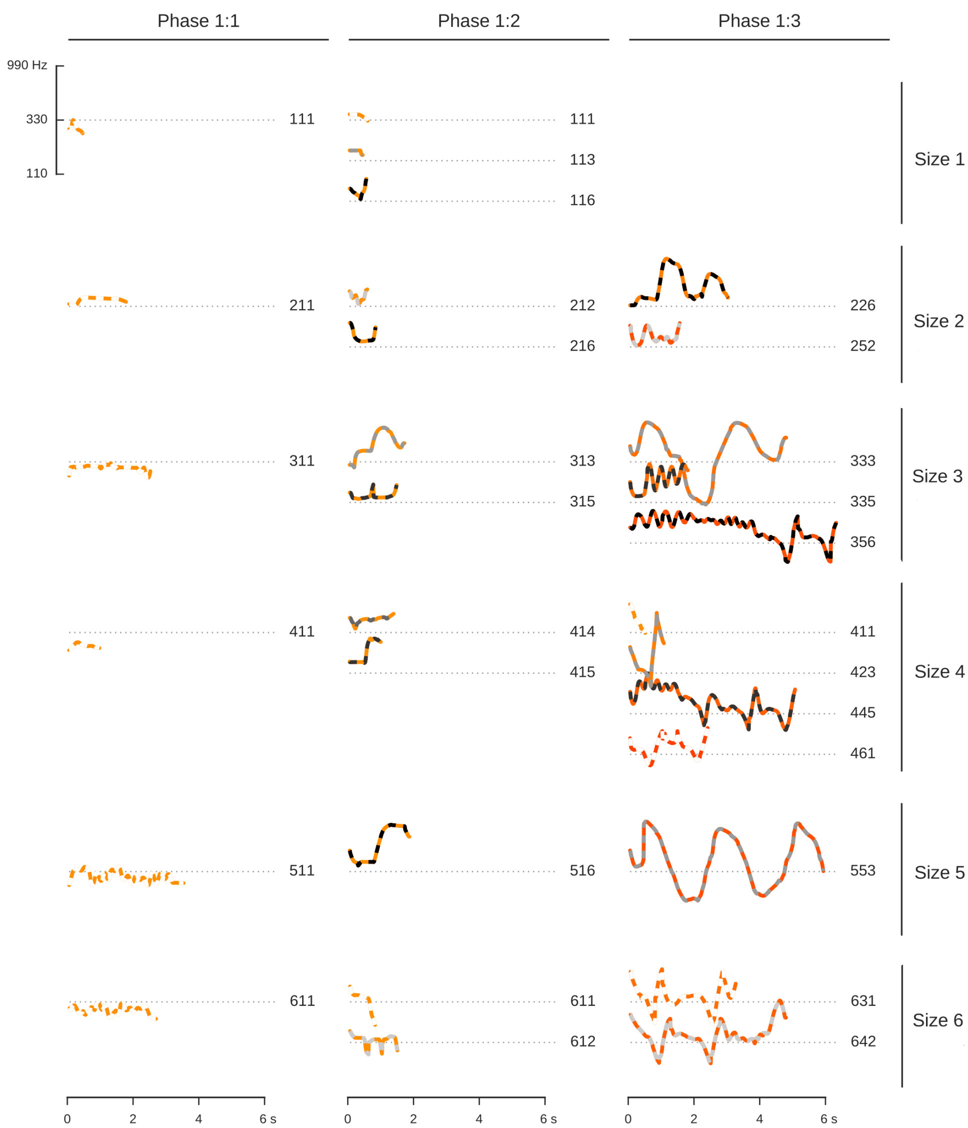

Figure 6.5: The entire signal repertoire of one participant in all three phases. Time is on the x-axis and pitch is on the y-axis. The colour of the stripes in the pitch tracks represents the colours of the squares they represent. Signals for smaller squares are higher in the diagram. The numbers by each pitch track are the values given to each individual meaning, derived from their size, shade and colour. Signals produced in phase 1:3 have visibly more movement than in the other two phases.

phase created an issue especially because all the information in the meaning space couldn't be encoded in a static signal. Despite the failure, I think HMMs are still a very worthwhile method for measuring structure in continuous signals.

6.6.4 Signal Recognition Task

Overall, participants were slightly worse at recognising their own signals in this experiment than in Experiment 2, however, they recognised their signals with a mean of 56% correct (s.d. 13%), again, chance level was 25%. Using a linear model, I tested whether participants improved in their performance throughout the experiment, as they did in Experiment 2, but found no correlation ($F(1,73) = 1.39$, $p = 0.24$). Success stayed constant across phases around the 56% mean. The lack of improvement as participants became more experienced was probably because the meaning space expanded so rapidly with each phase, making the recognition task much more difficult.

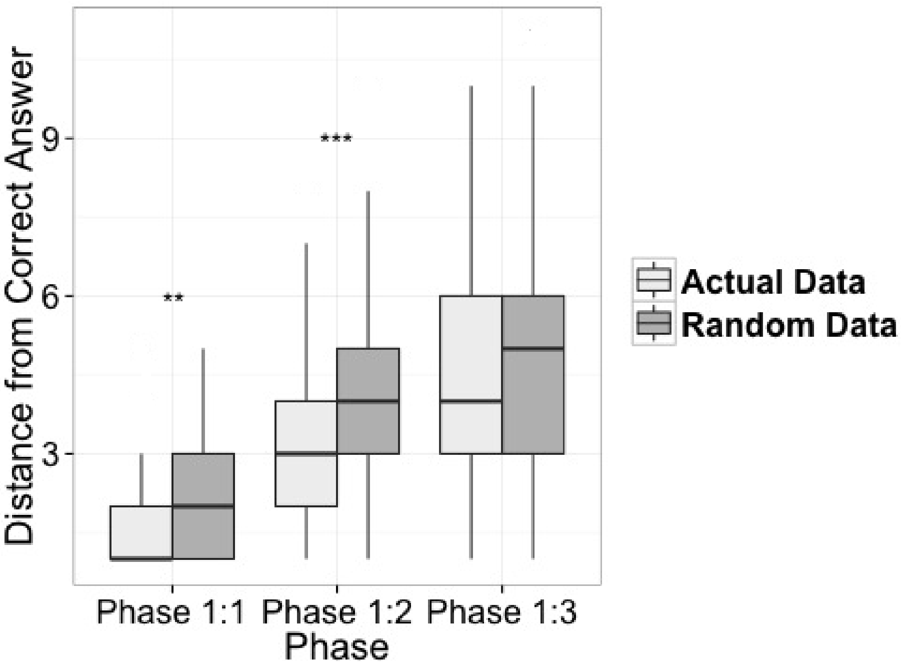

Figure 6.6: A graph showing the distance from the correct answer participants were in each phase when choosing incorrectly in the signal recognition task.

Again, when participants were incorrect, I was able to measure the distance between their answer and the correct answer. I did this in the same way as I did in Experiment 2. Using a mixed effects linear model, and controlling for participant number as a random effect and square number as a fixed effect, I found that with incorrect choices produced across phases, participants were closer to the correct square (3.3 steps away, s.d. 2) than if they had chosen at random (4 steps away, s.d. 2.1) ($\chi^2(1) = 22.4$, $p < 0.001$) (see figure 6.6), the difference between actual and random result was significant within each phase as well.

In later phases, incorrect distances were higher because of the bigger meaning space. Therefore, 4 meanings chosen at random would have a much bigger mean distance between them with bigger meaning spaces. As a result, comparison of distance from the correct answer between phases is not indicative of participants having problems. However, the difference in effect size when comparing the actual data with random data might indicate within which phase the signals were the most iconic (as choosing meanings close to the correct meaning when incorrect indicates use of diagrammatic iconicity). The effect size for the comparison between the actual data and the random data in phase 1:3 was smaller ($r = 0.13$) than in the other two phases ($r = 0.24$), suggesting that in phase 1:3 there was less potential to rely on iconic strategies, even though all effect sizes remain quite low.

I used a linear model to test if musical proficiency predicted performance in the signal recognition task, and, as in Experiment 2, found that it did not ($F(1,23) = 0.03, p = 0.28$).

6.6.5 Post-experimental questionnaire

In Experiment 3, no one self reported not having a strategy. Generally, participants in Experiment 3 reported the experiment to be more difficult than participants in Experiment 2. In phase 1:1, participants encoded size directly with pitch or duration (80% self reported). Participants tended to stick with the same strategy for size, but developed strategies on top of that to cope with the different shade elements, and by phase 1:3, 56% of participants self-reported as using a strategy which relied on movement, patterns or pattern frequencies.

Responses to the colour categorisation part of the questionnaire were very variable, ranging from 2-6 categories over the 6 squares, with a mean value of 4.2 categories, though most categories included the word orange, such as“light orange”,“dark orange”, “red orange”, “sunset orange”, “blood orange”, but people also labelled the darkest shade “red”. There was no interaction between the number of categories that participants separated the squares into and how well they did in phase 1:3 (which was the only phase to use different shades of orange) (F(1, 16)= 1.56, $p = 0.23$).

6.7 Discussion

In this chapter, I set out to experimentally investigate two scenarios:

1) When signal and meaning spaces have the same number of dimensions, this facilitates the emergence of iconic signals.

2) When the number of meaning dimensions is larger than the number of signal dimensions, this facilitates the emergence of combinatorial structure.

These two scenarios are intrinsically linked, as they are both tied up in the dimensionality of the signal space being used. This line of reasoning makes the experimental evidence relevant to the differences seen between how combinatorial structure emerges in different modalities, as different language modalities offer different levels of dimensionality (as argued in Chapter 3), and therefore different levels at which they can iconically represent the world.

In both experiments, I found correlation between the structure of signal repertoires and the structure of the meaning space, indicating a prevalence of diagrammatic iconicity. This was particularly marked when signal and meaning spaces had the same number of dimensions. I also found evidence for more movement in signals in phases where there was a mismatch between signal and meaning spaces, suggesting a departure from diagrammatic iconicity to a possibly more structural signalling system. Signals were also longer in later phases in Experiment 3, which perhaps points to more sequential encoding. Lewis & Frank (2016) previously showed that longer word forms are associated with meanings with more complexity, and signal duration has also been used as a measure for complexity in experimental studies such as Roberts et al. (2015).

Phases with matching dimensionalities produced signals that were more predictable given a participant's entire repertoire, than signals produced within mismatching phases. This, again, is indicative of the mismatching phases producing signals with more movement, as less complex static signals will be easier to predict.

I also found that in matching phases, when participants were incorrect, they were more likely to choose meanings that were closer to the correct meaning than if they had chosen at random, again suggesting a reliance on diagrammatic iconic strategies.

The above results provide evidence for the first mechanism, that matching dimensionalities produce signals with diagrammatic iconicity. They also show that more movement and complexity was present when meaning dimensions outnumbered signal dimensions. However, exactly how I can characterise this movement remains unclear. One possibility is that the movement in the signals is iconic, for instance, representing the stripes of meanings in Experiment 3. However, the post-experimental questionnaires do not support this narrative. It is clear from post-experimental questionnaires that participants often used structural strategies, but when they are present, they are not indicative of combinatorial structure as I define it in Chapter 2. That is, the building blocks are not meaningless but correspond to dimensions in the meaning space. However, there is very little flexibility in the way signal elements can be combined in the experiment (a result of my experimental design), and parts of the signals/meanings cannot occur in isolation. In this respect, the structure is neither combinatorial nor compositional but something in between, and possibly something that could be reanalysed by speakers to be combinatorial structure through the mechanisms proposed by Goldin-Meadow & McNeill (1999). Investigating what might cause this reanalysis to happen would make a good starting point for future experimental work, perhaps having participants creating signals for bigger and less structured meaning spaces to get rid of the inhibiting effects of iconicity.

6.8 Conclusion

I have shown that the dimensionality of a signal space will affect the emergence of structure and iconicity: the more closely the dimensionalities of the signal and meaning space correspond, the easier it is to use iconic structure. If there is no good correspondence, we see the first steps towards structure (either combinatorial or compositional). These results may help explain why combinatorial structure in spoken language is more prevalent and necessary than it is in the manual modality that has more signal dimensions. With more dimensions, I have argued this will not only affect how quickly the signal space gets overcrowded, but also how similar the topology of the signal space is to the meaning spaces we have in the real world with many dimensions. This theoretically follows from work claiming that the manual modality is better suited to iconically representing the world.

My results are also important for researchers conducting artificial language experiments with signal space proxies. The topology of the signal space being used could have significant effects on the iconicity and structure that emerges in the experiment. Importantly, understanding these effects, as I have attempted to do here, will then allow us to separate them from other effects under investigation elsewhere in the literature, such as expressivity and learnability.

CHAPTER 7

Signal-Meaning Mapping[1]

This chapter explores what happens when a meaning space presents a challenge for signal-meaning mappings beyond simply having more dimensions. I explore different types of iconicity that come from different available signal-meaning mappings and how these different types of iconicity affect the structure that emerges in signals.

7.1 Introduction

In the last chapter, I discussed how dimensionality mismatches can disrupt diagrammatic iconicity and the effect that this has on the emergence of structure in signals. However, as I discussed in the introduction, there is not only one type of iconicity. Different types of iconicity may affect the emergence of structure in different ways. Iconicity may stop combinatorial structure from emerging because there is no real need to change from an iconic system that has high referential efficiency. However, diagrammatic iconicity is only useful for ease of reference if one knows more than one signal in the signal repertoire. A diagrammatic system is also especially sensitive to crowdedness in a semantic space, as all signals need to exist on the same signal dimension. Further, signal and meaning spaces need to be able to map on to one another, which is facilitated by them having the same number of dimensions.

In this chapter, I present an experiment that disrupts the ability to use diagrammatic iconicity, defined by the presence of a correlation between the structure of a

[1]Part of this chapter is published as a CogSci proceedings paper (Little et al., 2015).

signal repertoire and the structure of a semantic repertoire. I do this not by affecting the dimensionality of the signal or meaning space, as I did in Chapter 6, but by changing the meaning space to not be "ordered" in a way that lends itself well to representation using a continuous signal space. With categorical dimensions, rather than continuous ones, it is much more difficult to create a generalisable system of diagrammatic iconicity. However, whether categorical dimensions provoke strategies such as combinatorial structure, or other forms of iconicity, such as imagic or lexical iconicity, is something that has not been explored experimentally.

7.2 Types of iconicity and the Emergence of Structure

In Chapter 2, I explored the notion that combinatorial structure may result from a process of conventionalisation. Once a signal has been grounded with the use of iconicity, it then loses its iconic features as the result of repeated interactions. This was shown experimentally by Garrod et al. (2007). Once it has lost its iconicity (or the iconicity has become dormant), elements of the signal open themselves up to be potentially reanalysed as meaningless building blocks. One thing I explore in this chapter is what factors are likely to affect this process. The notion of having "elements" of a signal implies signals are made up from multiple elements in the first instance, either in sequence or in parallel. Multi-part signals, when they are iconic, are perhaps more likely to be closer to imagic iconicity, where signals resemble or have something in common with their referents in some way, than diagrammatic iconicity. This is because the more elements an iconic signal has, the more elements of a meaning it is likely to encode, creating a more accurate iconic representation. Perhaps, then, signals with high levels of imagic iconicity, rather than diagrammatic iconicity, are more likely to provoke the emergence of structure because they are more likely to include multiple elements that can be interpreted once they have lost iconicity.

Previously (as already covered in this thesis) experimental work has explored how disruption of iconicity will elicit structure in signals (Roberts & Galantucci, 2012; Roberts et al., 2015; Verhoef et al., 2015a). However, these studies often treat iconicity as an individual entity that is on or off, i.e. comparing imagic iconicity with an example where no mapping is possible, rather than comparing the effects of different types of iconicity, which is the focus of the current chapter.

7.3 Experiment 4

In the last chapter, I demonstrated that meanings with continuous dimensions facilitate diagrammatic iconicity with a continuous signalling space. This mapping between a continuous signal space and a continuous meaning space is roughly analogous to spatial systems in sign languages where continuous meaning spaces (distances or angles) are described using a continuous signal space (e.g. respective distances or angles of articulators). In the experiment in this chapter, I explore what happens with more discrete meanings where diagrammatic iconicity is not so useful.

7.4 Hypotheses

My hypothesis for this experiment is that mapping a continuous signal dimension to unordered or discrete meaning dimensions will cause:

1. Less diagrammatic iconicity than when continuous meaning dimensions are used

2. More evidence of structure

7.5 Methods

Experiment 4 is again a signal creation experiment and has an identical set up to Experiment 2. However, the meaning space is different. The meanings in this experiment were designed to be discrete and have no ordering ("Discrete condition"). The results of this experiment are compared to Experiment 2 as a control condition (the "Continuous condition")[2].

7.5.1 Participants

Participants in the discrete condition were recruited at the VUB in Brussels. 30 participants took part in the experiment, 17 male, 13 female. Participants had an average age of 23 ($sd = 3.45$). Participants in the continuous condition are summarised in the

[2] I wrote in Chapter 6 about some problems I had with Experiment 2. Nevertheless, Experiment 4 uses Experiment 2 as its control, rather than (the better designed) Experiment 3 because I ran Experiment 4 before I had designed Experiment 3. However, I do not see this as a problem as the issues with Experiment 2 do not interfere with the issue I'm interested in in this chapter (i.e. continuous vs. discrete meanings). Further, it would have been unfeasible to have a discrete meaning space the size of the meaning space in Experiment 3.

methods for Experiment 2. As in Experiment 2, I asked participants to list the languages they speak with the level of proficiency. No participants reported knowledge of any sign languages. Participants also self-reported their musical proficiency (on a scale of 1-5).

7.5.2 Signals

Signals were produced in the same way as they were in Experiment 2 with the Leap Motion controller. There were 2 possible signal dimensions; moving a hand in the vertical dimension affected volume and moving a hand in the horizontal dimension affected pitch. The signal space was either 1 or 2 dimensional depending on phase (summarised below).

7.5.3 Conditions

The only thing which differs between condition is the meaning space. The meaning space in both conditions consisted of a set of squares that differed along 1 or 2 dimensions.

1. **Continuous Condition**
   Experiment 2 is treated as a condition which is labelled the continuous condition. In this phase, squares differed in either size and shade of grey.

2. **Discrete Condition**
   Squares differed in colour, though the colours were ordered to not be perceived as having order, and in their texture (see Fig. 7.1).

7.5.4 Procedure

Participants were instructed on how to create signals using the Leap Motion. They had time to practice creating signals. As in Experiment 2, participants knew from the beginning that they would be doing signal recognition tasks so they understood the benefit of creating distinct signals.

There were three phases of the experiment with each phase consisting of a practice round and an experimental round. There was no difference between practice rounds and experimental rounds. Only the data from the experimental round was used in the

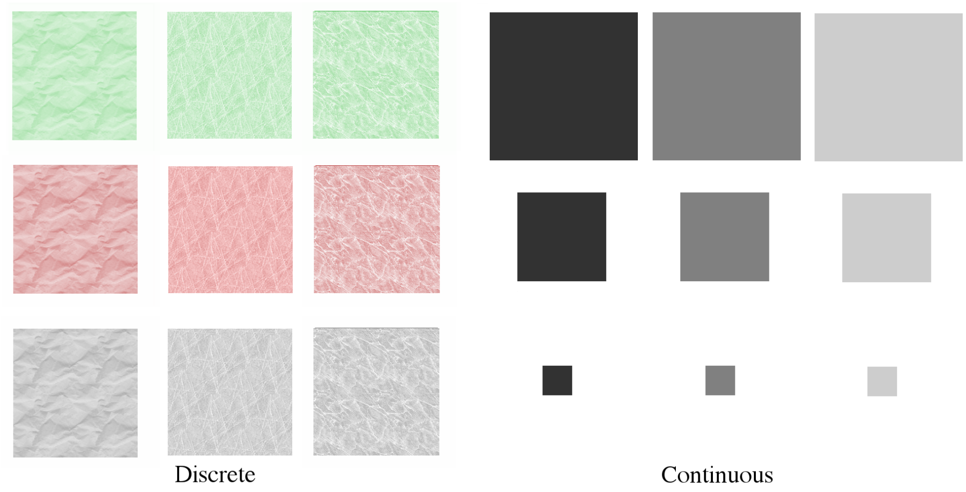

Figure 7.1: The meaning spaces used in phases 1:2 and 2:2 in the Discrete condition and in the Continuous condition respectively.

analysis. Each practice and experimental round consisted of a signal creation task and a signal recognition task.

7.5.5 Phases

In all phases, participants saw the entire meaning space before beginning. In the signal creation task, they were presented with squares one by one in a random order and recorded signals using the leap motion. They manually started and stopped recording themselves using on-screen buttons. They could play back their signals and record them again if they were not happy. Participants were explicitly told which signal dimension(s) they were manipulating.

Phase 1:1

In the first phase, participants were asked to create signals for a meaning space with 5 squares which only differed in one dimension (size or colour depending on condition). In this phase, participants could only manipulate the signal with one signal dimension, which was counterbalanced by randomly assigning participants to start with either pitch or volume. Which signal dimension the participant started with was later controlled for in the analysis.

Phases 1:2

In phase 1:2, participants described a two-dimensional meaning space with the squares differing in both size and shade in the continuous condition, or both colour and texture in the discrete condition (Fig. 7.1). They used the same one-dimensional signal space used in phase 1:1 (see fig. 7.2).

Phases 2:2

In phase 2:2, participants were to describe the same two-dimensional meaning space as in phase 1:2, but this time with a two-dimensional signalling space where the signals differed in both pitch and volume (Fig. 6.1).

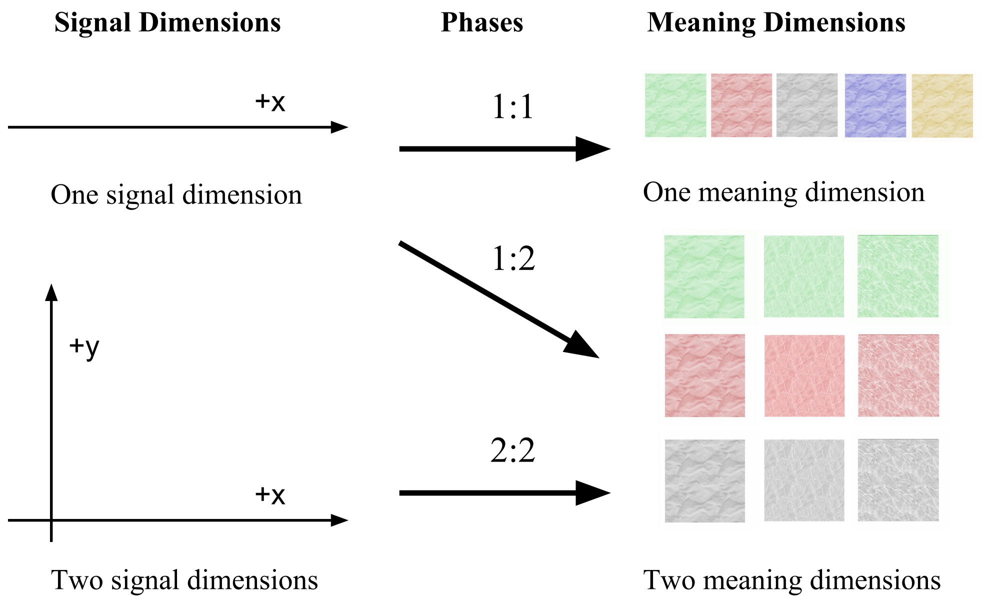

Figure 7.2: The mapping between signal space dimensionality and meaning space dimensionality in each phase using the meanings from the Discrete condition (the Continuous condition had the same mappings and is represented in Chapter 6).

Phase Counterbalancing

As in Experiment 2, participants were randomly assigned to do either phase 1:2 or 2:2 immediately after phase 1:1.

7.5.6 Signal Recognition task

After each signal creation task, participants did a signal recognition task. They heard one of their signals, and were asked to identify its referent from an array of four randomly generated possibilities (taken from the possible meanings in that phase). They had immediate feedback about the correct answer. Their performance in this task was recorded for use in the analysis.

7.5.7 Post-experimental questionnaire

After the experiment, each participant completed a questionnaire. Questions asked about what strategies were adopted in each phase of the experiment. The questions asked explicitly whether participants had strategies and what they were. Answers were free form.

7.5.8 Results

7.5.8.1 Post-experimental questionnaire

Post-experimental questionnaire results from Experiment 2 are reported in Chapter 6. Participants in the discrete condition were more likely to attempt forms of iconicity that were not the diagrammatic tendencies displayed in Experiment 2 (e.g. using pitch, volume and duration directly to encode the continuous meaning dimensions). In the Discrete condition, some participants associated the different colours with emotions or objects in the world like “a beating heart” for red, ”the sound of a cricket” for green or “the waves of the sea” for blue, and then made signals which corresponded to these things. Participants in the Discrete condition were much more likely to report strategies that relied on the use of patterns, speeds or frequencies of repeated elements than in the Continuous condition.

7.5.9 Signal Recognition Task

Which condition participants were in had an effect on how well participants performed in the signal recognition task. Participants were significantly better at the recognition tasks if they were in the Discrete condition (M=82% correct), than if they were in the Continuous condition (M=66% correct) (t(52.7) = -3, $p < 0.005$). Though, as discussed in the results in Chapter 6, the mistakes made by participants in the Continuous condition were somewhat easy to predict by how close the correct and chosen meanings were in the semantic space. This was more difficult to measure in the Discrete condition as distance between meanings was not as easy to measure because of their unordered nature.

The order in which participants received phases 1:2 and 2:2, and which signal dimension they started with (pitch or volume), did not reliably predict participants recognition of their signals. If a participant scored at chance level on the signal recognition task (N=1), they were disqualified from the rest of the analysis.

7.5.10 Signal Creation Task

Standard deviations (SDs) of signal coordinate values were used to measure the amount of movement in a signal. Following the hypotheses in section 7.4, I would expect signals in the Continuous condition to have smaller SDs than in the Discrete condition because one-to-one mappings between continuous signals and continuous meanings allows signals to encode all of their information in their static pitch or volume. Further, the post-experimental questionnaires suggest strategies that use a lot more movement.

As predicted, in the Discrete condition, there was a tendency for SDs of signal trajectories to be bigger than in the Continuous condition (23% on average, see figure 7.3 to see how this translated to the signals that were actually produced). I used a linear mixed effects analysis and controlled for participant and meanings as random effects, and whether they started with volume or pitch, phase number and order of phases as fixed effects. I found that this effect was not significant in either the standard deviations of the first dimension seen ($\chi^2(1) = 1.4$, $p = 0.24$) or the second dimension seen ($\chi^2(1) = 0.9$, $p = 0.34$). I also did not find an affect on SD regarding whether the number of meaning dimensions matched the number of signal dimensions in the Discrete condition. This effect was found for the Continuous condition and is reported in Chapter 6. Again, using a mixed liner model and controlling for participant, meaning (random effects) and phase order, phase number, and staring signal dimension (fixed effects), I did not find a significant effect ($\chi^2(1) = 2.17$, $p = 0.14$), however SDs were higher in the mismatch phases by around 12%.

I also tested whether duration of signals was predicted by condition. Again, running a linear mixed effects model and controlling for participant number, meaning as random effects, and starting signal dimension, phase number and order of phases as fixed effects, I found that signals were not more likely to be longer in one condition

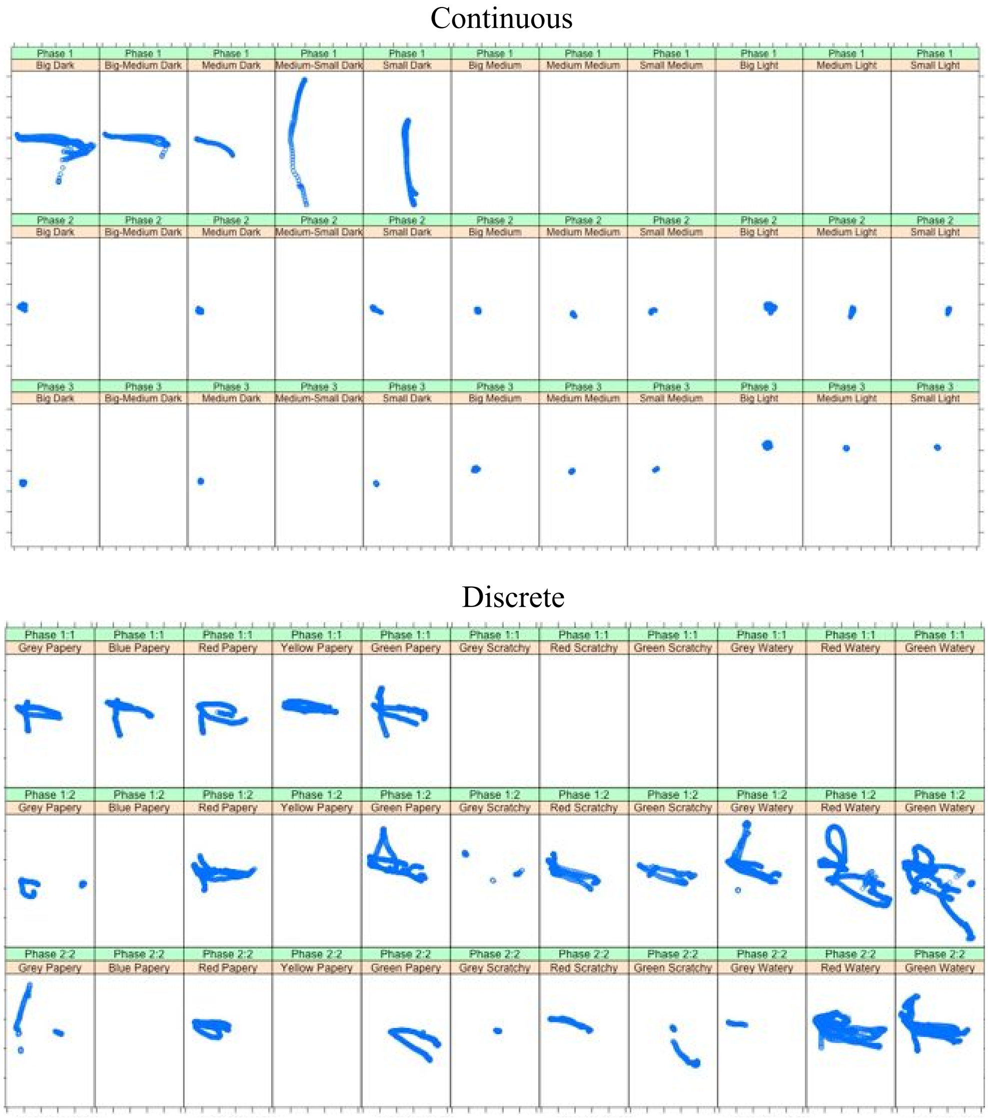

Figure 7.3: An example of a repertoire from each of the conditions (continuous and discrete). Each row of signals is the last production from each phase. The meaning that each signal refers to is described in the orange bands (though the primary function of this figure is giving a global idea of the amount of movement between conditions). The y axis represents hand position on the y-axis of the articulation space. The represents hand position on the y-axis of the articulation space.

than the other ($\chi^2(1) = 0.8$, $p = 0.38$). Though in the Discrete condition signals were a little longer than in the Continuous condition by around 12%.

7.5.11 Predictability of signal trajectories

I found an effect of condition on how predictable signals were within a repertoire (as defined in Chapter 4), using a linear mixed effects analysis and controlling for the same random and fixed effects as above ($\chi^2(1) = 12.3$, $p < 0.001$). The Continuous condition showed more predictability than the Discrete condition.

7.6 Discussion and Conclusion

In this chapter, I set out to experimentally investigate whether making the meaning space discrete and unordered will affect the use of diagrammatic iconicity and affect the use of structure within signals. I hypothesised that with a discrete meaning space, signals will be more complex due to having to adopt imagic iconicity rather than diagrammatic iconicity, and this in turn would generate more movement and structure in signals.

I did indeed find that participants self-reported using strategies involving patterns and frequencies. They also used more forms of iconicity that encoded features of the squares or some property or external object that they had associated with that colour or texture. This was in contrast to the ubiquitous use of diagrammatic iconicity used in the Continuous condition. There was also quantitative evidence that signals were more complex in the discrete condition as their trajectory SDs were bigger (though this was not significant) and the signal repertoires were less predictable.

One of the most interesting results to come out of this experiment is that participants were better at recognising their signals in the Discrete condition, than in the Continuous condition. This wasn't necessarily related to any of the hypotheses, but does speak to some of the theories in the existing literature on iconicity. Monaghan et al. (2012) have hypothesised that sound symbolism is beneficial to category learning, but not beneficial for learning individual words. This is an effect of sound symbolism making words more similar within categories, making them more easily confused within the category, but more recognisable as a group when being contrasted with words outside of that category. Something similar may be at work with the signals produced here. While one might think that having a system of diagrammatic iconicity would make a signalling system more intuitive, and perhaps easier to be productive with, it may also make signals more confusable. The pressure against diagrammatically iconic systems in the discrete condition may have pushed participants to make more exaggerated differences between their signals within their chosen strategies, which then aided them in the recognition task.

7.6.1 Modality Effects

Iconicity in signed languages can be both imagic and/or diagrammatic. This is also true for spoken language where phenomena such as onomatopoeia can be characterised as imagic iconicity, but diagrammatic iconicity is also possible. For example, continuous dimensions in phonemes, such as vowel quality, have been found to encode meaning dimensions such as size in ideophones (Dingemanse, 2012). It is difficult to quantify or predict whether one type of iconicity is more likely in one modality than another, and so it is difficult to say that these effects might be more dominant in one modality than the other. However, since the sign modality has more signal dimensions (as argued in Chapter 3), it may be easier to be iconic overall using both types of iconicity. This may mean that the effects seen in both conditions here, that are driven by iconicity, may be stronger in signed languages for signs with different types of iconicity.

7.6.2 Types of iconicity and Conventionalisation

One thing I was also eager to see with this experiment was evidence of behaviour that may affect conventionalisation. Further to imagic iconicity provoking more structure, it is possible that imagic iconicity may be more susceptible to conventionalisation than diagrammatic iconicity. Imagic iconicity necessarily requires icons to be abstracted from their objects in the first instance because only some features will be represented perhaps facilitating conventionalisation. Diagrammatic iconicity, on the other hand, does not necessarily require abstraction, but parallelism. The whole signal system needs to be maintained in order for the iconicity to be retained and be useful in any way, creating a pressure against conventionalisation. With imagic iconicity, the loss of iconicity, or iconic elements, would only affect the sign in question.

It is important to note that while the above paragraph hypothesises that conventionalisation is more likely with imagic than diagrammatic iconicity, I have not tested these specific hypothesese in the current experiment because there is no adoption, propagation or interchange of signals. Previous experimental work has shown that conventionalisation is a process that requires more than one person in repeated interaction Garrod et al. (2007) and in the current experiment, only one participant repeats signals for themselves. As a result, I would like to use this section to talk a little about how the different types of iconicity produced in the experiment manifested and how this might influence conventionalisation drawing on examples from the existing literature.

In signed languages, signs tend to lose their iconicity over time as a result of conventionalisation (Frishberg, 1975). However, some signs lose their iconicity at different rates (or not at all). Consider signs for spatial referencing and directions. In sign languages, these tend to often be very iconic compared to spoken language (Emmorey, 2002). These signs often rely on indexical or diagrammatic iconicity to make reference to directions or locations in relation to the speaker or another fixed point. It is difficult to imagine these signs becoming completely arbitrary because the iconicity is so useful for spatial referencing. The function of the signs would also not benefit from categorisation, which is something that results from a process of conventionalisation. However, signs for more categorical meanings (such as animals) do tend to become less iconic as a result of conventionalisation, for example, the ASL sign for BIRD once had flapping wings, but no longer does (Frishberg, 1975).

In the Discrete condition in the current experiment, there were more individual differences in the strategies used by participants than in the Continuous condition. There was also more variability in those strategies. For example, only one participant connected red to hearts, only one participant connected blue to waves and only one participant connected green to grass. In contrast, in the Continuous condition similar strategies were being used for similar reasons. For example having bigger squares correspond to longer signals and having lighter shades correspond to higher pitch/quieter volume. Adding communication between more than one person in this experiment will necessarily produce signals that are less idiosyncratic to an individual in both conditions, but especially in the Discrete condition because the signals are so much more idiosyncratic to start with. This will also massively affect the process of conventionalisation because participants will need to agree on signals, establishing iconic conventions before signals can even start to lose iconicity through further conventionalisation.

In addition to affecting conventionalisation, having more discrete, imagic and idiosyncratic forms of iconicity can also affect the amount of movement in signals because of the amount of information being encoded in the signals. This, in turn, affects the number of units which may be conventionalised and reanalysed in the emergence of combinatorial structure.

In conclusion, the type of iconicity possible may facilitate conventionalisation and the emergence of structure in different ways. However, in order to investigate how conventionalisation would affect different types of iconicity, longer experiments would need to be carried out, with communication and/or transmission added in order to replicate processes seen in experiments such as Garrod et al. (2007). However, in the absence of further experimental work, the findings here provide evidence that combinatorial structure could have emerged as a result of the nature of mappings between signal and meaning spaces and how they combine. Meaning dimensions with less intuitive ordering may facilitate imagic iconicity and combinatorial structure, where more ordered structure in a meaning space can facilitate diagrammatic iconicity and something more similar to compositional structure. In order to shift the focus back to combinatorial structure, the rest of the experiments in this thesis use a meaning space that is designed to have less internal structure than the meaning spaces presented so far.

CHAPTER 8

Communication[1]

This chapter presents experiments that look explicitly at whether an expanding meaning space will affect the emergence of combinatorial structure through pressures of discrimination, as well as looking at the effect of adding communication. Further, these experiments investigate whether conventionalisation will occur in speech-like signals in the same way that it does in gestural and graphical signals.

8.1 Introduction

8.1.1 Discrimination

So far, this thesis has explored how size or dimensionality of signal spaces might affect how quickly a signal space gets crowded. However, the experiments have either had no meanings (Experiment 1), or have presented the entire meaning space to a participant in the first instance (Experiment 2, 3 and 4). In these latter experiments, while the meaning space corresponded to the signal space in different ways in different conditions (with some conditions having more meanings than others), there was no condition that directly tested the effects of the number of meanings in isolation from other confounds relating to mapping. The experiments in the current chapter have expanding meaning spaces. An expanding meaning space should have the same effect as a small

[1]Part of this chapter is published in Interaction Studies (Little et al., 2017a) and some further playback experiments have been published in the CogSci proceedings (Little et al., 2017c)

signal space on crowding in the signal space. The increasing number of meanings should increase pressure for combinatorial structure emerging because of signal space crowding (which is discussed in Chapter 2).

It may not seem ecologically valid to have a meaning space expanding. Our ancestors did not live in an expanding world. However, in order to bootstrap a communication system it is necessary to start with a small number of signals before expanding the system to refer to other meanings. This is also true for language acquisition; children start with a small number of signals, typically pertaining to meanings which are salient to them in their immediate environment, before the wider world becomes relevant to them as they grow. It is therefore not ecologically invalid to model this growing number of signals with a growing number of meanings.

8.1.2 Communication

One of the most debated issues in language evolution is whether structure in language is primarily the result of pressure for communication or iterated learning transmission. Many studies have conditions comparing pressures for expressivity (communication) to learning (e.g Carr et al., 2016; Fay et al., 2010), in order to tease apart what aspects of structure are explained by which kind of transmission. They often find that a combination of both (or the combination of pressures for both expressivity and learnability) creates the optimal environment in which structure emerges (Kirby et al., 2015). As far as continuous signal spaces looking at the emergence of combinatorial structure go, there is some work which uses, for instance, Galantucci's (2005) stylus paradigm in communication games (Galantucci, 2005; Roberts & Galantucci, 2012; Roberts et al., 2015), as well as work using an iterated learning experiment (Del Giudice, 2012). Further, slide whistles have been used in iterated learning experiments (Verhoef et al., 2014, 2015a, Chapter 5 of this thesis) and also communication games (Fullerton, 2011). This work has found that structure emerges as both the result of communication and learning. However, no work on signals to date has made a direct comparison. This will be discussed further in the "further work" section in the concluding chapter. Instead of comparing communication with iterated learning, here I will compare an individual signal creation experiment with a communication experiment. While there is a pressure for expressivity in Experiments 1-4, there is no second individual to communicate with. So there is a pressure for expressivity (having a signal be distinct in order to be useful for communication), however, this is not the only thing which communication creates pressures for. Knowledge that a signal is meant for someone other than oneself will drive signals to be less idiosyncratic. By less idiosyncratic I mean that there will be more of a pressure for transparency in both the structure and the iconicity in the signals that doesn't necessarily exist when someone is producing signals for themselves. However, previous work looking at processes of conventionalisation demonstrates that communication between more than one individual is what leads signals to become less iconic over time, which doesn't happen when a lone individual repeats a signal over and over without a communication element (Garrod et al., 2007).

8.1.3 Conventionalisation

One often cited feature of language is that it is arbitrary (Hockett, 1960; De Saussure, 1983). This is largely true, though it is now accepted that language is a lot more iconic than linguists previously thought, especially in signed languages (Perniss et al., 2010), but also in spoken languages (Blasi et al., 2016). However, iconicity is one way in which language could have initially bootstrapped itself as a communication system (Imai & Kita, 2014). If a signal is similar to its referent in some way, it will be easier for language users to establish a signal-meaning mapping. However, there is very little direct evidence available from real world languages about how language initially bootstrapped itself, especially spoken languages. As a result, experimental studies have been used by researchers in the field of language evolution to investigate the effects of interaction and transmission on levels of iconicity and symbolism in signals. Specifically, studies have concentrated on how we could have got from iconic beginnings to an arbitrary system via processes of conventionalisation.

One of the main methods for investigating the process of conventionalisation has been the field of experimental semiotics (see Galantucci & Garrod, 2011, for a review). This started as far back as Brennan & Clark (1996), where participants communicated about different tangrams. Tangrams are arrangements made up from 7 flat shapes. They found that after repeated interactions, the statements about the tangram arrangements became more simplified as participants started to use elements of the original tangram descriptions as "short-hand". This simplification of originally complex forms, is the hallmark of conventionalisation.

Other studies that used graphical signs to investigate conventionalisation include Theisen et al. (2010), which used a pictionary style paradigm in a communication task. They showed that over the course of the communication game, drawings became less iconic. One of the contributing factors to minimise production effort in this experiment was an incentive for participants to have as many successful communicative interactions as possible within a constrained time period. A slightly different approach was demonstrated by Caldwell & Smith (2012), which had "replacement microsocieties", where they had a constant turnover of naïve participants who contributed to signs becoming simpler and more abstract. One of the driving forces for signs becoming simpler in this experiment was that participants could interrupt the production of a signal once they were sure what it was, meaning signals never had to be more complex than they needed to be. Concurrent feedback, such as interruption, was also found to drive conventionalisation in conditions in Healey et al. (2007) and Garrod et al. (2007).

There have also been several studies that have used gestural experiments to investigate whether conventionalisation happens through interaction to get from iconic pantomime-like gestures to more arbitrary language-like symbolic gestures. Namboodiripad et al. (2016) used a communication game in the lab to get participants to repeatedly communicate scenes to one another and were able to measure hallmarks of conventionalisation over the course of the experiment. Duration of gestures and the size of the space used for the gestures was reduced, as was the amount of complexity within a gesture, though they did not measure iconicity directly. Motamedi et al. (2016) also investigated conventionalisation in gestural languages, but focused on the effect of transmission rather than interaction, looking at how signs changed in an iterated transmission chain, where participants' signs were learnt from those output of a previous participant pair. This study found that gestures developed from pantomimes to less complex, more arbitrary signs with compositional structure.

Real world data can also contribute to our knowledge of conventionalisation processes. There is diachronic evidence of some signs in American Sign Language (ASL) losing complexity and iconicity (Frishberg, 1975; Schlehofer, 2016). This process of conventionalisation may be what facilitates the emergence of the first combinatorial phonology-like elements in emerging sign languages as discussed in Chapter 3 (Sandler et al., 2011). Roberts & Galantucci (2012) showed in their experiment that phonology-like combinatorial structure was present when signals were more conventionalised (less iconic), though it is not clear if this happened over the course of the experiment, as their analysis did not include data across time. As already discussed in Chapter 2, experimental work has also looked at whether iconicity inhibits the emergence of combinatatorial structure (Roberts et al., 2015; Verhoef et al., 2015a, and work in this thesis Experiments 2-4). It is possible, then, that conventionalisation (and the loss of iconicity) is the thing that triggers the emergence of combinatorial structure in language.

8.2 Hypotheses

Above, I have provided discussion on how communication might affect signals. There are two alternative hypotheses: the first hypothesis is that communication (in comparison to signals produced by a lone individual) will drive signals to be more transparent in their iconicity. The second hypothesis is that communication will drive signals to be less iconic due to the process of conventionalisation as has been shown in previous experimental work (Garrod et al., 2007).

Further, the way communication systems become bootstrapped necessarily means that the number of signals expands because, a) signals can't all be bootstrapped at once and b) the number of items to be communicated may expand as more things become relevant over time. As I discussed throughout chapter 2, an increase in the number of signals will cause signal space crowding which will cause pressures for combinatorial structure. In the current chapter, I will also test a third hypothesis that an expanding meaning space will increase complexity in signals in order to maintain distinctions between signals.

8.3 Experiments

There are 2 signal creation Experiments presented in this chapter: Experiment 5, an individual creation experiment similar to those presented in chapters 5 and 6 where participants both produce and recognise their own signals, and Experiment 6, a communication game experiment where participants take it in turns to produce and recognise the signals of their partner.

These experiments use the same meaning space and signal space, and are largely comparable in their structure. These experiments were designed in a similar way to aid comparison. I have avoided calling them conditions here as conditions of an experiment should only ideally differ in one aspect, which is the aspect under investigation. Unfortunately, methodological problems that arose when piloting the communication experiment meant that changes needed to be made to the communication experiment.

8.3.1 Garrod et al. (2007)

Further to comparison between the experiments presented in this chapter, I will also compare the results to Garrod et al. (2007). Garrod et al. (2007) is an experimental semiotics study that investigated, among other things, the effect that communication will have on the process of conventionalisation of graphical signals. Garrod et al. (2007) had 3 conditions. In one condition, one participant repeatedly drew items for an imaginary audience (no feedback). In another, one participant drew items but were given feedback from a partner. In the final condition, two participants took it in turns to draw items for each other with ongoing feedback. The study measured complexity in the images throughout the task, as well as the levels of iconicity in the drawings. They measured iconicity with the rate at which naïve listeners could match the drawings with their intended referents after the experiment. Garrod et al. (2007) showed that knowledge of early interactions in the communication condition of the experiment improved naïve listeners' ability to match drawings produced later in the experiment with their referents, but without this knowledge naïve listeners were a lot worse at matching later drawings. This indicates that the images were becoming less iconic without the knowledge of how signals were grounded. Getting naïve listeners to match signs with referents is now a common method used in experimental semiotics to measure iconicity. If naïve listeners can pair signals with their intended meanings, then those signals can be said to be iconic. Garrod et al. (2007) also found that complexity in the images dropped throughout the communication condition, as it did in Brennan & Clark (1996). However, in the individual condition, with no communication partner, the drawings increased in complexity.

In this chapter I will present experiments which are similar to the individual condition, with one participant repeatedly producing signals for themselves (Experiment 5), and the communication condition, where participants take it in turns to produce signals for each other (Experiment 6). Drawing (as used in Garrod et al., 2007) is a highly iconic modality, and in this regard is more similar to the manual modality than speech. Since my signals are continuous, auditory and constrained in their iconicity, it will be interesting to discover what will happen with more speech-like signals. This is the primary motivation for comparing my results to that of Garrod et al. (2007). I also conduct a playback experiment with naïve listeners (Experiment 7) in order to compare my results with theirs.

It is worth noting here that my experiments differ from those is Garrod et al. (2007)

in a number of ways other than just the modality. For instance, the meaning space is very different. In Garrod et al. (2007), the meanings were places, people, TV programmes, objects and abstract concepts. In my experiments, the meaning space is a much simpler set of shapes. In Garrod's experiment in the individual condition participants were not asked to recognise their own drawings (presumably because they would behave at ceiling on this task). However, in my experiment, I do ask individuals to recognise their own signals. This gives me comparable measures of recognition accuracy from within the experiment, which Garrod did not have. It also creates a pressure for expressivity in the individual condition (as well as the communication condition) which allows me to isolate the effects caused by the communication of two people, rather than the process of communication itself (as individuals are effectively communicating with themselves).

There was also a difference between how we define feedback in our experiments. In Garrod et al. (2007), in the comparable communication task, participants were able to draw on each others' signals after they had been produced and transmitted. Sometimes this could be just a tick to indicate they got it correct, which would be analogous to the feedback in my experiment which is simply being told if you are right or wrong. However, in Garrod et al. (2007) a lot more flexibility was allowed in the form that the feedback could take which may have played a role in the process of conventionalistion. Such flexible feedback is a lot more difficult to provide with the Leap Motion signals. Garrod et al. (2007) also ran another experiment with concurrent feedback (feedback allowed throughout signal production) which did significantly aid the conventionalisation of signals.

8.3.2 Signals

Participants again created signals using the Leap Motion hand-tracking device that affected the pitch of signals along the horizontal dimension, the same as in Experiment 3. No other dimensions were used. Signals were created by pressing a button on screen to start recording, and another to stop the recording. There was no time limit on how long signals could be. Participants could play back their signals, and rerecord their signals if they were unhappy with them.

8.3.3 Meanings

Experiments 2-4 were concerned with how signals mapped to meaning spaces. As a result, the meaning spaces were very structured in order to understand exactly what part of the meaning space structure was causing the results. However, as I discussed in Chapter 6, this caused some problems with interpreting the structure that appeared in those experiments as structure in signals was directly mapped to structure in the meaning space. Though the emergence of structure on any level is interesting, and combinatorial structure can emerge from compositional structure, this thesis has its primary focus in the emergence of combinatorial structure. Accordingly, the meanings used in this chapter are designed to not be structured in order to get away from having structure emerge that mirrors the structure of the meaning space. The meanings all had three features: shape, colour and texture. No two meanings had any features in common, for example, in figure 8.1, only the left shape had the features blue, circle and stripey, and only the cross had the features grey, cross and wavy lines. There were 15 meanings in the experiment (see firgure 8.2).

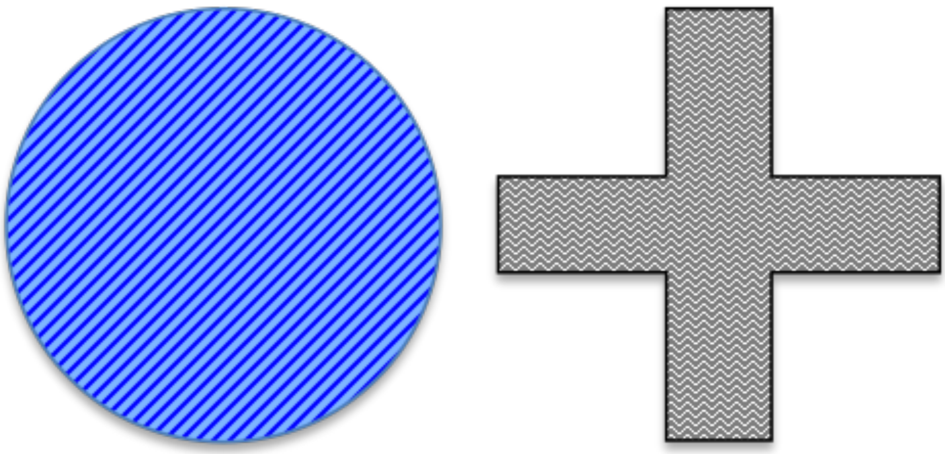

Figure 8.1: Two meanings with different shapes, colours and textures.

8.3.4 Experiment 5 (Individual)

8.3.4.1 Participants

24 participants (17 female, 7 male, average age 21 ± 1.3) took part in the individual experiment and were paid €5 for the 30 minutes it took to complete the experiment.

Participants were recruited at the Vrije Universiteit Brussel. I collected data on their musical proficiency (self reported on a 5 point scale from ”none” to ”expert”), as well as the languages they speak and their proficiency levels in each. All participants gave informed consent to have their data recorded and knew they could leave the experiment at any point.

8.3.4.2 Procedure

Participants were given clear instructions about the structure of the experiment. They were explicitly told how many phases there were, and how the phases were structured. They were also told how to use the leap motion by simply moving their hand either left or right to manipulate the pitch of the signal. They got to try this out before the experiment began. Participants knew from the beginning that they would have to recognise their own signals in each round.

8.3.4.3 Phases

Participants created signals in three phases (see figure 8.2). In the first phase, they created signals for 5 meanings, chosen at random from the pool of 15. In phase 2, they created signals for all of the meanings they had already seen, plus 5 more, making 10 in total. In phase 3 they created signals for all 15 meanings.

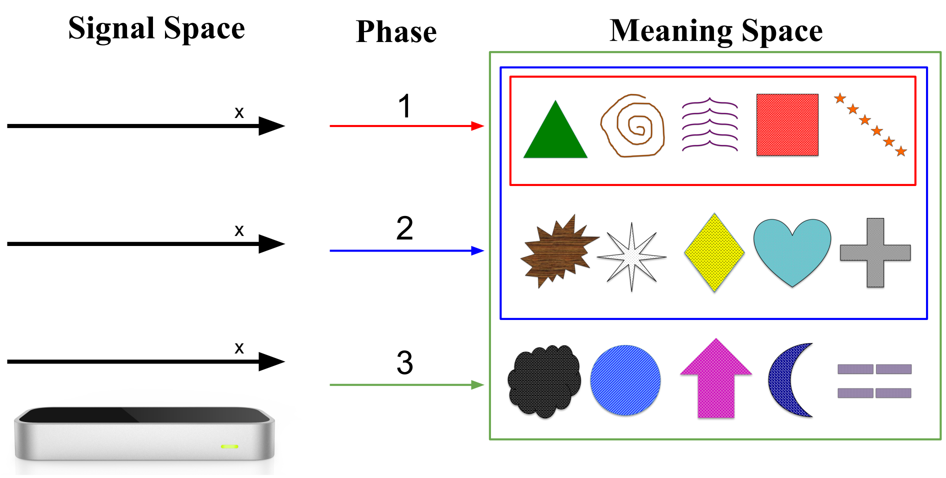

Figure 8.2: The shapes used as meanings in the experiment in the 3 phases in the individual experiment, with the meaning space increasing by 5 with each phase.

8.3.4.4 Signal Creation Task

Before the signal creation task, participants were told that they would see images that they need to create signals for. They were explicitly told they should make sure they remember the signals as they would be asked to recognise them during the experiment. This introduction screen also displayed the whole meaning space for that phase, so participants knew which meanings were in a phase before they began creating signals.

Meanings were presented one after another in a random order and participants created a signal for each one by pressing a “record” button to start, and a “stop” button to finish. Participants could play signals back and rerecord them if they were not happy.

8.3.4.5 Signal Recognition Task

Once participants had created signals for all meanings within a phase, they were given a signal recognition task. They heard each of their signals, one after the other, and were asked to identify the meaning the signal referred to from an array of 4 choices. The array included the correct meaning, and 3 other meanings taken randomly from the subset of meanings used within the current phase. Participants were given feedback on the correct answer immediately after each response.

8.3.4.6 Practice and Experimental Round

Participants completed the signal creation task and the signal recognition task twice for each phase. The first time was framed as a “practice round”, and existed so that the participant could get used to the structure of the experiment and how to use the apparatus. Only data from the experimental round was used in the analysis of the experiment.

8.3.4.7 Post-experimental questionnaire

After the experiment, participants completed a post-experimental questionnaire. It asked about the specific strategies participants used to generate signals, and whether they felt their strategies changed at all throughout the course of the experiment.

8.3.5 Experiment 6 (communication)

8.3.5.1 Participants

32 participants (27 female, 5 male, average age 20.9 ± 2.8) took part in 16 pairs in the communication experiment. In this experiment participants were paid €10 for 1 hour. Participants were recruited at the Vrije Universiteit Brussel.

8.3.6 Key differences from Experiment 5

As mentioned previously, this experiment had several differences extra to the experimental manipulation (1). These differences nearly all derive from the communication task being much more difficult than the individual task.

(1) The main difference is that Experiment 6 has 2 participants communicating, rather than 1 participant communication with themselves

(2) The second biggest difference is how the meaning space expands. In Experiment 5, the meaning space expanded with 5 meanings at a time in batches. In Experiment 6 the meaning space expanded by 2 meanings at a time.

(3) The third difference is when the meaning space expansion happens. In Experiment 5, the meaning space expands every phase (defined as one full practice round and one full experimental round). In Experiment 6, the meaning space expands only when every meaning has been "established". I define what I mean by established and this process in detail below.

(4) This experiment took longer to complete, and participants were therefore paid more money. Further, no participants finished the experiment in Experiment 6, but all paticipants finished it in Experiment 5.

(5) Participants in the communication task were given an incentive to finish the task more quickly (a €20 voucher).

8.3.7 Procedure

Participants were given clear instructions about the structure of the experiment. They were told they would be playing a communication game. Again, they were told how to use the sensor and given an opportunity to practice making signals before the experiment began. They were told how a turn worked. They were told that they would be given feedback about their success after each turn, and that they would take it in turns to produce signals. If participants had not finished the experiment after 50 minutes, then the experiment would automatically end. They were told this also.

Participants also knew that the experiment would progress more quickly the more successful they were (the specifics of this mechanism are explained below, though the specifics of this were not explained to the participants in detail). Participants were also given an incentive to try to finish the experiment quickly. They were told that the pair of participants who did the experiment the fastest would win a €20 voucher.

Two participants took it in turns to produce and receive signals. They could not see each other. One turn in the experiment involved the producer creating a signal for a given meaning. They saw the meaning on the screen and had to record a signal for that meaning using the leap motion. They pressed a "record" button and stopped the recording when they were done using the same button. They could play back the signal and re-record it if they were not happy. They then pressed "send" and the signal was sent to the receiver. The receiver listened to the signal by pressing a button when they were ready, they then had to choose what meaning they thought the signal referred to from an array of up to four meanings. They could not select a meaning before they had listened to the signal. After each turn, both participants were given feedback about whether their communication was successful (i.e. whether the meaning the signal was produced for was the meaning the receiver selected). They also received feedback about what specific meaning the producer was communicating, and the meaning the receiver chose. They then swapped roles and started again.

While a participant waited for the other participant to finish creating or recognising a signal, they were given a text screen instructing them to wait.

As in the individual experiment, the communication experiment also had an expanding meaning space by phase. However, the meaning space only expanded by 2 meanings at a time (rather than 5 in the individual experiment) and the experiment only continued to the next phase once the participants had agreed on signals for existing meanings. Ideally, the meaning space should have expanded at the same rate as the individual experiment. However, participants found the communication game much more difficult than I had anticipated when they were given the meanings in batches of 5. As a result, I designed a system where the meaning space expanded in line with their success in the experiment, in order to not overwhelm the participants with too many meanings at once.

Participants started with 2 meanings, chosen at random from the 15 possible meanings. The array in the recognition task in this first "phase" was constrained to these 2 meanings until there were more meanings and then the array was always 4. Once meanings had been communicated correctly twice in a row, they were considered "established" meanings. If an established meaning was communicated incorrectly, it would lose its established status. Once all meanings in a phase were established (all meanings seen until that point), then the meaning space expanded by 2 more meanings, starting a new phase. Since there were 15 meanings, the meaning space expanded only 7 times, making 8 phases in total.

At first, which meaning the pair were to communicate in each interaction was presented at random. However, once the meaning space expanded once, meanings were chosen for interactions with a probability determined by whether it was an established meaning or not. Meanings were chosen with a 45% probability if they were established, and the remaining 55% of the time the meanings were either newly introduced meanings, or meanings that had recently been communicated unsuccessfully. This mechanism was in place because if all meanings had the same probability of appearing throughout the experiment, the experiment would take far too long. Unestablished meanings needed to have a reasonable frequency in order to become established so that the experiment could progress.

Once all meanings were established, the experiment finished automatically. If participants did not achieve established signals for all 15 meanings before 50 minutes, they were stopped and their interactions and signals were recorded up until that point.

The signal data used in the analysis of this experiment was taken from signals once they had become established, in order to make them more comparable with the signals created in the experimental rounds in the individual experiment.

Participants completed a post-experimental questionnaire, as in the individual experiment.

8.3.7.1 Signal Measures for Conventionalisation

Garrod et al. (2007) had a metric for perimetric complexity in the drawings in their study that used measures of the inside and outside perimeter of the drawing, and the amount of ink used (Pelli et al., 2006). This measure was used as an approximation of the amount of information represented in a graphic. For the analysis in this chapter, I equate the duration and the amount of movement in a signal as being somewhat analogous to Garrod's complexity measure as these metrics should represent the amount of "auditory ink" that has been used. If signals are conventionalising, then the amount of complexity should decrease. This is backed up in pictionary-style tasks, such as

Garrod et al. (2007), and also in real-world languages like ASL, where size of signing space and the duration of signals has been found to be lost over time for some signs (Namboodiripad et al., 2016). I again measured the amount of movement in a signal with the standard deviation of the trajectory of x-axis coordinates in each signal and the duration with the number of data frames in a signal.

8.4 Results

The results will be presented in two parts: those results pertaining to the signals in Experiments 5 and 6, followed by the results pertaining to the signal recognition tasks, both within Experiments 5 and 6.

8.4.1 Signals

8.4.1.1 Movement in signals

To investigate what affected the amount of movement in signals, I conducted a linear mixed effects analysis, with standard deviation of signal coordinates as the dependent variable and how early in the experiment a signal was produced (phase number) and experiment as fixed effects. I had participant number and meaning as random effects, as well as by-participant and by-item random slopes for the effect of both time produced and experiment. I then conducted likelihood ratio tests of our model against a null model without the effect in question in order to obtain p-values. I found that the amount of movement in a signal was influenced by whether it was created in the individual or communication experiment ($\chi^2(1) = 6.9$, $p < 0.009$), with signals from the individual experiment having standard deviations that were lower, indicating less movement in the signals (see figure 8.3). However, how early in the experiment participants produced the signals did not significantly affect movement in the signals ($\chi^2(1) = 0.13$, $p = 0.25$), and there was no significant interaction between experiment and time produced.

8.4.1.2 Duration of signals

I conducted a similar linear mixed effects analysis as with the standard deviation values above, to investigate the length of signals with the same random and fixed effects and random slopes. Signals produced in the communication experiment were longer

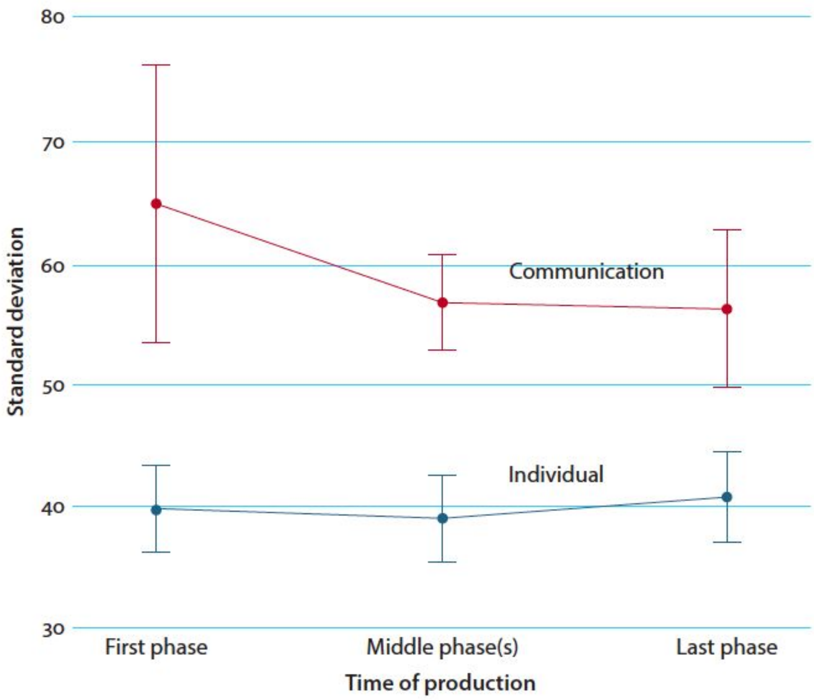

Figure 8.3: The standard deviation of coordinates within signals, indicating the amount of movement in signals, produced in both experiments, at the beginning, middle and end of the experiment. Here, "last phase" means phase 3 in the individual experiment, or the last phase that participants got to in the communication game.

than in the individual experiment, though this effect was not significant ($\chi^2(1) = 0.4$, $p = 0.52$). However, the time produced (phase number) did have a significant effect on the duration of signals ($\chi^2(1) = 4.4$, $p = 0.03$). As can be seen in figure 8.4, the duration went up throughout the experiment in both experiments, though this was more marked in the individual experiment.

8.4.1.3 Predictability of signals

I conducted a similar linear mixed effects analysis as above looking at predictability (as defined in section 4.5.3) with the same random and fixed effects and random slopes. Which experiment signals were produced in did not have an effect on the amount of predictability within signals ($\chi^2(1) = 0.02$, $p = 0.879$). There was an overall significant trend of production time but I broke down the model into experiments as the effect was being driven primarily by the individual experiment. How early in the experiment participants produced the signals significantly affected predictability in the

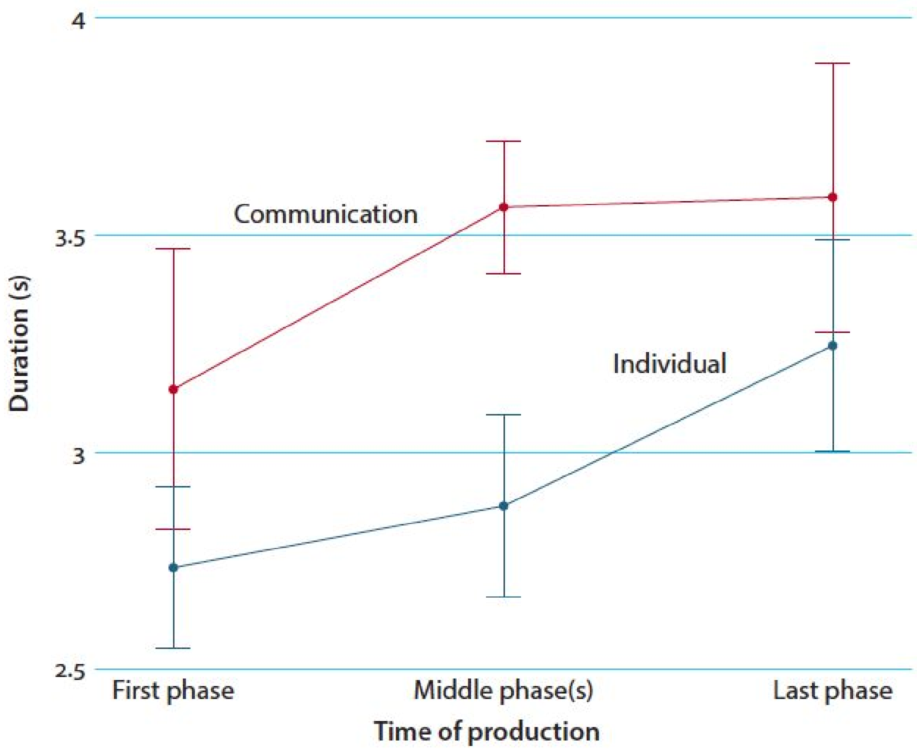

Figure 8.4: The durations of signals produced in both experiments, at the beginning, middle and end of the experiment. Here, "last phase" means phase 3 in the individual experiment, or the last phase which participants got to in the communication game.

signals in the individual experiment ($\chi^2(1) = 4.04$, $p = 0.04$), with a consistent loss of predictability (increase in complexity), as can be seen in figure 8.5, there is also an example of a repertoire from the individual experiment in figure 9.2 (next chapter). In the communication experiment there was no significant trend ($\chi^2(1) = 2$, $p = 0.16$).

8.4.2 Signal Recognition

8.4.3 Recognition of signals within the experiment

I conducted a linear mixed effects analysis to look at participant success throughout the experiments, with time produced and which experiment signals were produced in as fixed effects. I had meaning and participant (or pair) number as a random effect, as well as by-meaning random slopes for the effect of time produced. As above, I then conducted likelihood ratio tests of my model against a null model to obtain p-values. Which experiment signals were created in had a significant effect on participant

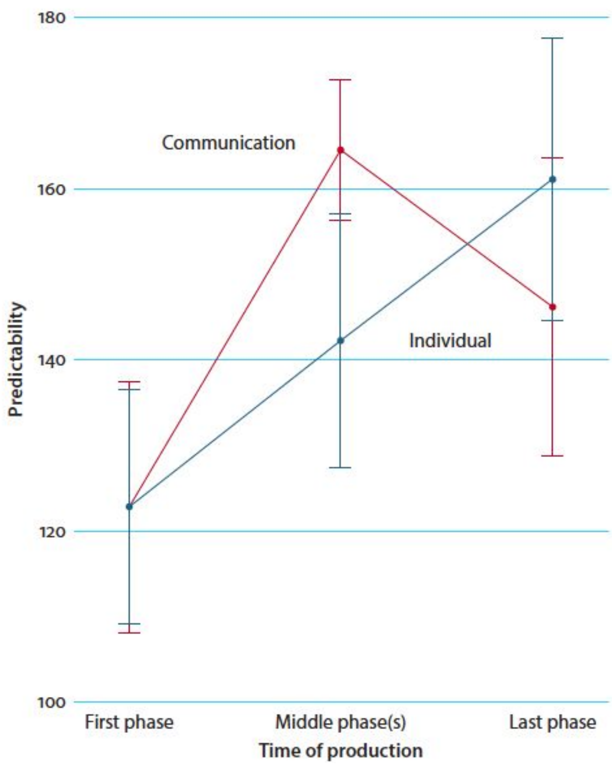

Figure 8.5: The predictability of signals produced in both experiments, at the beginning, middle and end of the experiment. Here, "last phase" means phase 3 in the individual experiment, or the last phase which participants got to in the communication game. Higher numbers here refer to lower predictability (or high complexity).

success within the experiment ($\chi^2(1) = 7.8$, $p = 0.005$), with participants being better in the individual experiment (85.6% correct) than in the communication experiment (74.4% correct). There was no significant effect of time produced on success during the experiment ($\chi^2(1) = 0.35$, $p = 0.55$). However, there was a significant interaction between experiment and time produced ($\chi^2(1) = 5$, $p = 0.02$). As can be seen in figure 8.6, in the individual experiment, participants got slightly better throughout the experiment. In the communication experiment, participants got worse.

Another measure of success within the communication experiment was how far participants got before their time ran out. As explained in the methods, whether participants got to the next phase was dictated by whether they had managed to establish signals for all of the meanings that were currently in the meaning space. As one would expect, some pairs were much better at the task than others, with some pairs only reaching the second phase of the experiment (4 meanings), and others doing much better (success of all pairs can be seen in figure 8.7). No pair managed to establish signals

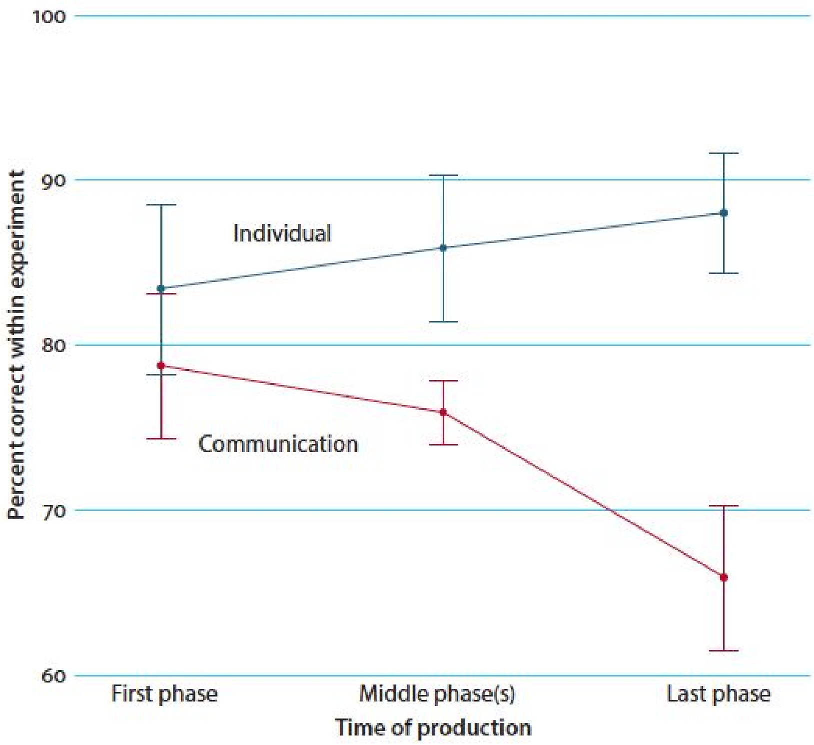

Figure 8.6: Scores of participants within the experiments, at the beginning, middle and end of the experiments. They are not cumulative, but a sample of responses from phases at the different periods. Again, here, "last phase" means phase 3 in the individual experiment, or the last phase which participants got to in the communication game.

for all 15 meanings, thus, nobody finished the experiment. All data used in the analysis throughout the paper was taken from signals made at the beginning or end (or middle) of the experiment in the communication experiment, rather than relying on using data from specific phases.

8.4.4 Post-experimental questionnaire

The questionnaire revealed that nearly all participants attempted to use iconic strategies throughout the experiment in both experiments. They were more likely to try and encode shape than any other feature.

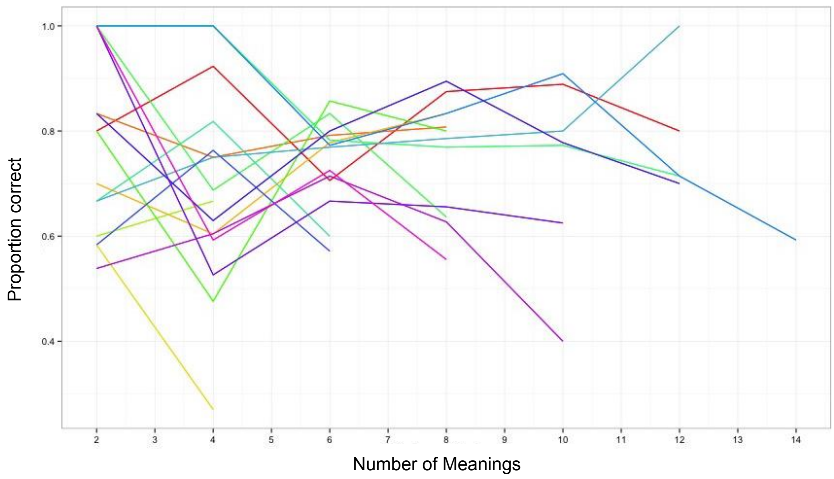

Figure 8.7: The success of participants throughout the communication experiment. The scores are not cumulative, but the percent of correct responses within each phase of the experiment, defined by the period before each meaning space expansion. Each pair is one line, and the length of the line illustrates how far that pair got within the 50 minute time limit.

8.5 Experiment 7 (playblack)

In experiments described in previous chapters, the meaning space was structured, making iconicity possible to measure from meanings and signals alone. That is, distance between meanings can be measured within structured meanings spaces, and the transformation distance between signals can also be measured. However, the experiments described in this chapter, because there was no structure within the meaning space, no meaningful measure for iconicity could be generated using data from the meanings and signals alone. In order to measure iconicity, I used naïve listeners to match signals they hear with a meaning from a set of 4 random possibilities. If naïve listeners can match signals with their intended referents without knowledge of a signal's development, then that signal can be said to be iconic.

8.5.1 Participants

391 naïve listeners were recruited on social media. Each listener was asked to listen to 15 mp3 signals each and asked to choose from an array of 4 possible meanings for each signal. Some listeners matched fewer than 15 as there were some signals left over

after all signals had been assigned to a possible experiment of 15 signals (see below for details).

8.5.2 Signals

Signals from the individual experiment were either taken from the first phase (the first 5 signals produced) or the third phase (the last instance of all 15 signals), so I could measure how iconicity was affected by repetition of signals throughout the experiment, or how iconicity was affected by signals being newly created within an already established system.

Signals from the communication experiment were either from the first phase (for the first 2 meanings after they had been communicated correctly twice in a row) or they were the last successful instance of signals produced in the last phase of the experiment that a pair saw, depending on how well they did in the experiment. As such, they included both signals which were present from the beginning, as well as signals which were newly established. Some pairs got to later phases than others as they were more successful at producing established meanings.

8.5.3 Procedure

Naïve listeners were sent to a webpage where they were redirected randomly to 1 of 46 possible signal sets. Signal sets were either made from signals produced early in the experiment, or signals produced late. Signal sets were composed from signals by more than one participant.

Naïve listeners were asked to turn on their speakers (or be wearing headphones). Listeners could remain anonymous, or they could enter their email address to be informed of the results of the experiment. Instructions read: ”please play each sound and choose the image which you think the sound refers to”. They could play each signal as many times as they liked. They then had to choose one of 4 possible signals. They could only choose one meaning for each sound. Data was recorded for whether the listener was correct, and if they were wrong, what they had clicked on instead.

8.5.4 Results

I binned responses by creating a mean value for correct matchings for each signal within each possible signal set. I then conducted a linear mixed effects analysis, with time produced (early or late) and experiment the signal was produced in as fixed effects. I had meaning as a random effect, as well as by-meaning random slopes for the effect of time of production and experiment. Again, I conducted likelihood ratio tests of our model against a null model. Experiment did not affect the amount of iconicity in the signals ($\chi^2(1) = 0.1$, $p = 0.74$), with overall levels of matching nearly exactly the same (around 35% in both experiments). How early in the experiments (5 and 6) participants produced the signals also did not significantly affect iconicity ($\chi^2(1) = 2.3$, $p = 0.13$). However, there was a significant interaction between experiment and time produced ($\chi^2(1) = 5.9$, $p = 0.015$).

As can be seen in figure 8.8, naïve listeners were much better at matching signals with their intended referents that were produced later in the communication experiment. However, in the individual experiment, the signals went down in their iconicity, though this difference was much less marked than in the communication experiment. Also, iconicity was not a predictor for how well participants recognised their own signals in Experiment 5, which was probably because participants were so good at recognising their own signals.

The playback experiment also enabled me to measure the iconicity of signals for specific meanings. Figures 8.9 and 8.10 show the iconicity of each signal as measured using naïve listeners. Some meanings lend themselves to iconicity better than others. The upwards pointing arrow is particularly strong in its iconicity, almost certainly because having a signal with rising pitch is an easy way to represent this meaning using the Leap Motion. Signals for pointy images were also easy to recognise, though some participants in the Experiment 6 did report having trouble differentiating the signals of their partners' for the pointy meanings.

Data was also collected on how many times listeners clicked on particular meanings in the playback experiment, regardless of what the correct meaning was. This data is useful to investigate whether some meanings are simply selected more often overall because they lend themselves particularly well with the iconicity that is possible using the Leap Motion signals. If this was true there should be a correlation between how often signals were clicked and how often they were correctly matched by naïve listeners. There was no correlation. Figure 8.11 plots the number of times meanings were selected in the playback experiment against the number of times listeners were correct.

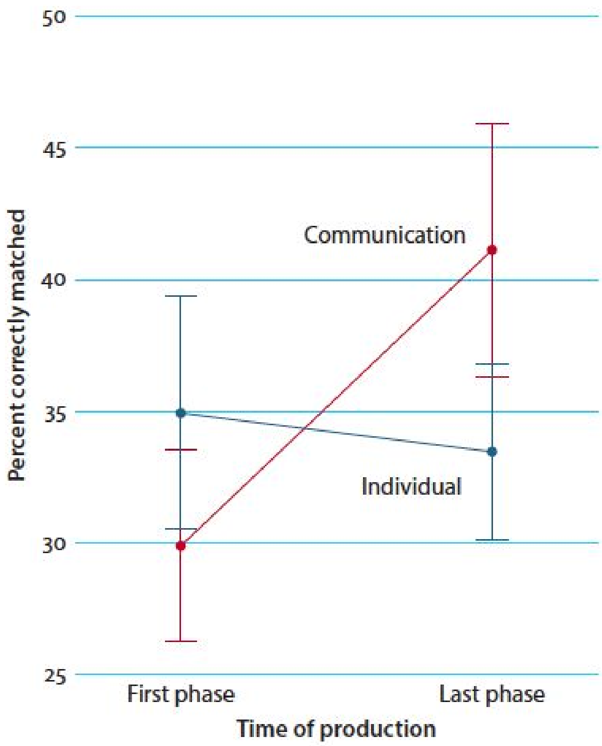

Figure 8.8: The percentage of signals currently matched with their meanings by naïve listeners. Both signals produced at the beginning and at the end of the experiment were tested. Here, ”last phase” means phase 3 in the individual experiment, or the last phase which participants got to in the communication game.

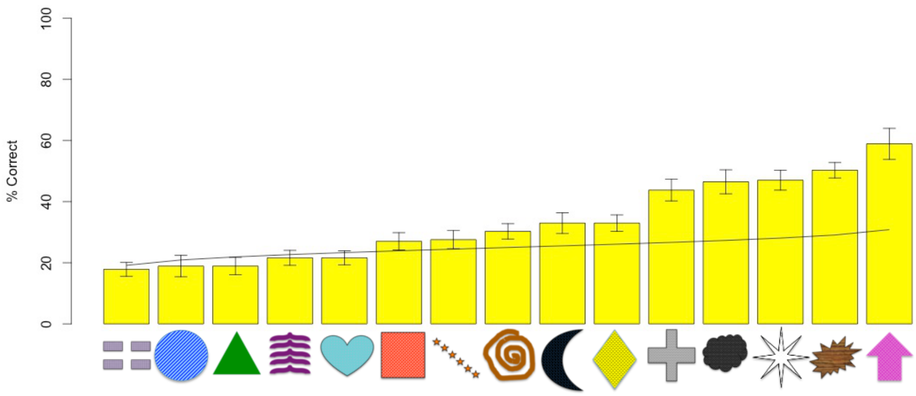

Figure 8.9: The percentage of correct responses from naïve listeners matching signals with their intended meanings. The graph shows data from the last phase of Experiment 5 with 95% error bars. The line represents what we would expect if matchers were behaving at chance level.

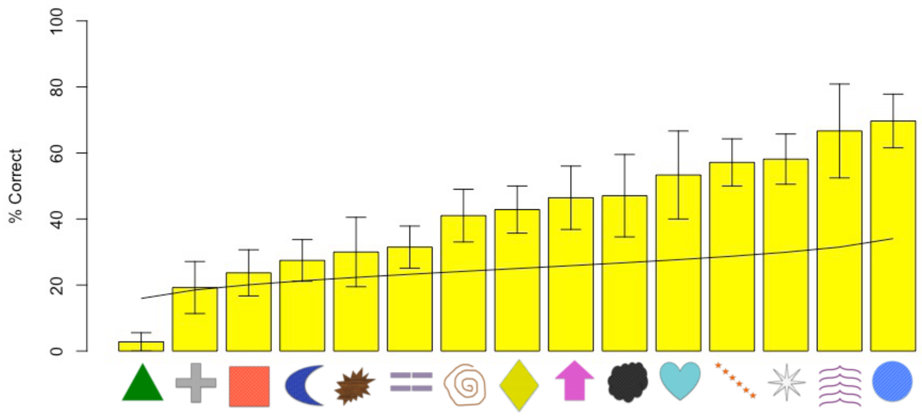

Figure 8.10: The percentage of correct responses from naïve listeners matching signals with their intended meanings. The graph shows data from the last instance of successful communication for each meaning in Experiment 6 with 95% error bars. The line represents what we would expect if listeners were behaving at chance level: it is not a straight line at 25%, because we need to take into account that the scores are ordered.

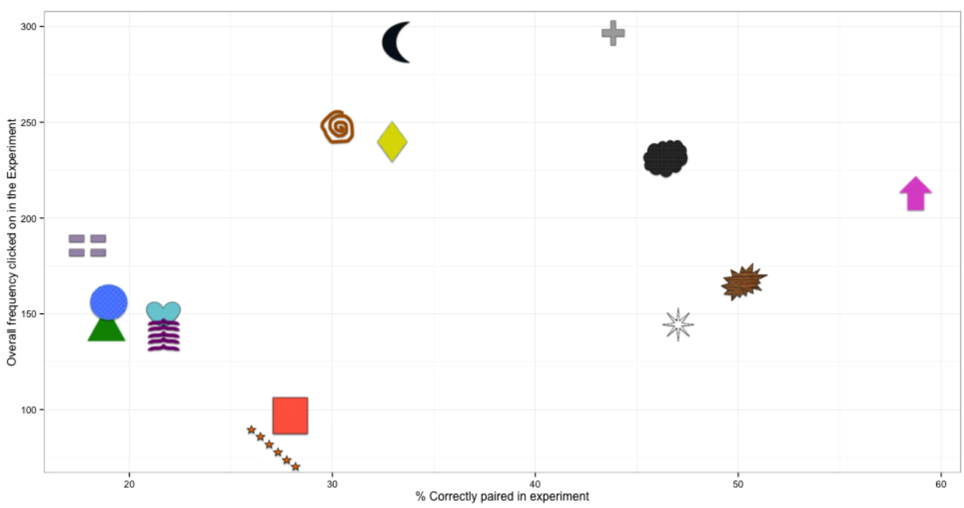

Figure 8.11: This graph plots the percentage that naïve listeners paired a meaning with its signal correctly out of the total number of times that meaning was the correct response against the total number of times a meaning was clicked on in the experiment, even when it was not the correct response. There was no significant correlation between these two variables.

8.6 Discussion and Further Work

In the experiments presented in this chapter, I measured complexity in the signals that were produced. This was to give some sense of whether signals were becoming more simplified throughout the experiment, from more complex iconic representations to a more abstract symbolic representation as would be expected from the results of previous work.

I have used several measures to quantify complexity in my signals (movement, duration and predictability). Using these measures, I found that signals were less complex in the individual experiment than in the communication experiment. This is in contrast to findings from Garrod et al. (2007), who found that pictures produced in their individual condition stayed complex throughout the experiment, and pictures produced in the communication condition reduced in complexity throughout, resulting in the images in the communication condition overall to be much less complex, the opposite of my finding. In my experiments, I found no effect of signals becoming more or less complex over time in the communication experiment, an effect that is likely to be due to my signalling paradigm having less available flexibility and affordance for iconicity than the pictionary paradigm. Signals in my paradigm are much more constrained in the forms they can take, which may mean they need to grow in complexity simply in order to differentiate between different meanings in the experiment.

Further to the above, under the hypothesis of Hockett (1960) that combinatorial structure emerged as a method to aid discrimination between signals, I might not expect the signals to become simpler in either the individual or communication experiments as further complexity is beneficial for the task of discrimination. The reason the drawings in the communication condition in Garrod et al. (2007) dropped in their complexity was possibly because their communication modality (drawing) was so much more flexible than the Leap Motion, allowing for more complexity as a starting point. With the signal-space being so much more restricted with the Leap Motion signals, participants started simple, and ran out of ways to generate distinctions between signals quite quickly. This again may have implications for the processes of conventionalisation (or the emergence of combinatorial structure) between languages in the real world, as the manual modality is arguably much more flexible than the spoken modality. This pressure for discrimination (or expressivity) is also often cited as important factor in the emergence of structure in artificial language experiments that use pre-discretised building blocks to form signals (e.g. in Kirby et al. (2008), Kirby et al. (2015), Carr et al. (2016)).

In my experiments I found that signals became more complex later in the individual experiment, or at least did not lose complexity, which is in line with the findings from Garrod et al. (2007). However, Garrod et al. (2007) hypothesise that their result is because, in the absence of feedback, participants encode more features in their signals later in the experiment as they think of more things they can include about the meanings they are communicating. The opportunity for this to happen in my experiments was relatively limited, as the meaning space was not so complex. Further, in post-experimental questionnaires participants usually only describe trying to encode one feature of the meanings (mostly shape). As a result, the pressure for discrimination, as described above, is a much more likely candidate for the growth in complexity seen in the individual experiment.

As mentioned in the introduction, Roberts & Galantucci (2012) found evidence that combinatorial structure was more likely to be present when signals were less iconic. If complexity growing in my signals is indicative of combinatorial structure and the result of reducing iconicity, signals should lose their iconicity as the experiment progresses. On the other hand, if iconicity grows, my growing complexity results may be more likely to be the result of pressures for discrimination from the growing meaning space.

I measured iconicity in the same way as Roberts & Galantucci (2012) and Garrod et al. (2007), by getting naïve listeners to match signals with their intended meanings. However, what I found was that signals became more iconic as the communication task progressed. Importantly, this is the opposite of Garrod's result, where naïve listeners who only saw drawings from the end of the experiment were worse at matching them to their correct referents than naïve listeners seeing the earlier drawings. Again, I can account for this result because of the fundamental differences between our paradigm and the drawings used by Garrod et al. (2007). It is much easier to be iconic with the more flexible drawing paradigm, allowing for more iconicity at the beginning, which can then be "lost". However, this does not account for the opposite trend that signals become more iconic I see in the communication experiment. It is possible that this happens because of participants becoming more accustomed to the communication game and to good strategies to use. Having another participant present with whom you are communicating may be driving the signals to be more iconic. Perhaps, the communication process causes signals to adapt to be more mutually intelligible. While signals produced by an individual for their own use may have a certain level of iconicity (at the levels found in the individual experiment), it is not necessarily true that this iconicity is transparent for naïve listeners. What makes a signal fit for communication may be iconicity that is less idiosyncratic. It may be that signals need to reach this level of transparent iconicity before they can be emancipated from their meanings in order to partake in the process of conventionalisation. In Perlman et al. (2015a), non-linguistic vocalisations also became more iconic over the course of a communication game possibly for similar reasons. Both vocalisations and the signals produced using the leap motion present a difficulty for producing transparently iconic signals. This difficulty is not present to the same extent when using gesture or drawing as modalities negating the need for an initial stage of negotiation. This explanation makes sense in the light of the signals not gaining iconicity in the individual experiment (see figure 8.8). It is also possible that transparent iconicity may require more complexity in order to be realised which would explain why signals became more complex.

I also found that participants within Experiments 5 and 6 were much more able to recognise signals in the individual experiment than in the communication experiment. In the individual experiment, no negotiation is needed to establish signals, which inevitably leads to higher scores. I also found that in the individual experiment, participants got slightly better throughout the experiment, despite the meaning space growing. This again could be because participants are simply becoming more used to the apparatus and task throughout the experiment. In the communication experiment, participants got worse, probably because the meaning space was growing, making the task more difficult, though as it only expanded by 2 meanings at a time, the effect of having new meanings to negotiate should not have affected the success rate throughout the experiment. However, new meanings competing iconically with old meanings could have affected success for both, and participants did self-report finding some meanings difficult to differentiate (e.g. the spiky brown shape and the white star). Previous artificial language experiments have demonstrated context effects on structure that comes out in these experiments (Winters et al., 2015). That is, signals only encode information that is relevant to successful communication, which may be different features depending on what other meanings are present. As the meaning space in the experiments presented here are designed to be unstructured, the effects of context are likely to be much less severe than experiments with structured meaning spaces. For example, if randomly selected meanings in the recognition task all had shared features this may produce different types of structure, and cause specific features to be encoded in signals. This wouldn't happen if all meanings had no shared features. I explore this idea a little more in the conclusion in the further work section.

I have not conducted an experiment exploring the affect of concurrent feedback, where participants could interrupt one another, because in my experiment, feedback only came after signals had been completed, transmitted and recognised. Previously, Healey et al. (2007) found that concurrent feedback in a task can be the driving force which makes representations more abstract and less iconic. Garrod et al. (2007) also ran a condition with concurrent feedback, and found that the loss of complexity proceeds faster with ongoing interaction throughout the production of drawings. Participants interrupting each other was also one of the driving forces for conventionalisation in Caldwell & Smith (2012). Concurrent feedback may be a worthwhile experiment to conduct using the Leap Moton paradigm. However, as signals are already so short (around 3 seconds), it may not provide much opportunity for interruption, and may in fact drive signals to be longer and more complex so that hearers can be more sure of their guess before interrupting.

Much of what I have covered in the discussion is speculation about my results. Concrete conclusions have largely eluded me, and therefore this work opens up a lot of avenues for future work to help answer some of the questions I have raised. Some potential future work is discussed in the concluding chapter of the thesis.

8.7 Conclusion

I have argued that conventionalisation, as a process for arbitrary forms emerging, may not work in the same way with different modalities. I found no evidence that signals in my experiments became more conventionalised (simpler and less iconic) through interaction or repetition. I hypothesise that pressures for discrimination due to an expanding meaning space are acting against the conventionalisation process causing signals to become more complex. Further, in modalities where iconicity is less available as a strategy, an initial stage of negotiation needs to occur in order to attain transparently iconicity before signals can become more arbitrary. More work needs to investigate whether concurrent feedback and repair may affect the process of conventionalisation using the Leap Motion. However, such feedback has not always been necessary to see the process of conventionalisation using other paradigms (Garrod et al., 2007). It is this disparity between my findings and those of other experiments using different mediums of communication, that suggests that the modality used to generate signals is an important factor involved in the dynamics of conventionalisation.

CHAPTER 9

Duration[1]

This chapter investigates the effects of signal duration on structure in signals. That is, whether having a time constraint on the production of a signal will affect the structure that emerges within signals. This is the first artificial language experiment explicitly investigating the effects of time constraints on signals.

9.1 Introduction

In the real world, there are several pressures that might restrict how long signals can be. Zipf was the first to observe that more frequent words tend to be shorter, indicating that there are pressures in language for speed, brevity and least effort (Zipf, 1949). Since then, Piantadosi et al. (2011) have shown that word lengths are strongly influenced by pressure for communicative efficiency, including considerations for the information content of words. Kanwal et al. (2017) have since done experimental work using an artificial language experiment that showed that participants used shorter forms with less information content when the context was more predictable. However, these results only play out when participants are under pressures for both efficiency and expressivity. These findings support the notion that language is subject to the principle of least effort.

Despite this clear evidence for a pressure for brevity, very little work has been done to investigate how pressures of duration affect the emergence of combinatorial units,

[1]Part of this chapter is published in the proceedings of EvoLangXI (Little et al., 2016).

and how quickly units are discretised and reused in shorter signals compared to longer ones. However, it is notable that some experimental studies that explicitly explore the emergence of combinatorial units constrain the duration of signals, while others do not. This is particularly troubling when there are good reasons to believe that the (maximum) duration of signals is very connected to structure on a combinatorial level.

9.1.1 What connects combinatorial structure and signal duration?

As mentioned in previous chapters, Hockett (1960) hypothesised that combinatorial structure comes about when one runs out of ways to make distinctions between a set of holistic signals. It would follow then, that with fewer possible signals, one must recombine signals sooner than otherwise. Furthermore, with fewer signal units, you will also run out of combined forms using only 2 units sooner than if you had more and so on. It would make sense, then, that languages with smaller phonemic inventories should have longer word forms to compensate. This is exactly what has been found by Nettle (1995, 1998), but only in a relatively small, though admittedly diverse, set of languages.

The above theory is neat, but it assumes discrete inventories as a starting point. It also assumes that the number of phonemes in an inventory is affecting the length the words can be, rather than vice versa. Here, I am instead interested in how the emergence of combinatorial structure is affected if signals start as holistic and continuous when there is a pressure for brevity.

On the one hand, pressure for shorter signals means a pressure for information density, efficiency in communication, and a pressure against the amount of redundancy in signals. Combinatorial structure (and compositional structure) is a great method to get a system to have information conveyed succinctly and efficiently with a very finite number of building blocks.

Furthermore, redundancy has the potential to detract attention from meaningful or repeated elements in a signal. As discussed in Chapter 3, Sandler (1996) made the argument that iconicity may be at the route of the persistence of phonologically redundant information (which is more likely to exist in longer signals). This redundant information will interfere with the process of analysing meaningless repeated chunks (as well as chunks that carry meaning), which will get in the way of structure emerging or being reanalysised at any level.

On the other hand, time constraints could inhibit the emergence of combinatorial structure. When there is more time to make signals longer, the time dimension can be used to counter crowding in a signal space on the other dimensions, rather than this happening on other signal dimensions in dissociable or indissociable ways. The time dimension is then the thing which facilitates the emergence of combinatorial structure rather than restricting it. Accordingly, it may be that longer signal durations may facilitate combinatorial structure to emerge, when the signal space is getting crowded through noise or other means.

9.2 Experiment 8

The experiment described in the current chapter is the first experiment to explicitly compare the effects of duration constrains on structure in continuous artificial signals. The results from this experiment are relevant in two domains:

1. **Duration effects on real world languages.**
   There may be pressures on how long utterances can be in real world, though as discussed in Chapter 3, these are not likely to be bound by modality.

2. **Duration effects on artificial language experiments.**
   In the current literature, some experiments have duration limits on signal production and some do not. It is therefore necessary to understand the effects which duration limits will have on signals that emerge in these experiments.

The second domain here I think is the most pertinent to language evolution research. We must aim to understand the effects of artefacts in our experimental designs, such as signal duration, in order to facilitate the interpretation of results. However, I will briefly expand on both domains below.

9.2.1 Time Constraints on Real World Languages

As already discussed, signals in the real world are subject to the principle of least effort, pressures for speed and pressures against redundancy. These will all push signals to be shorter. There are also physiological constraints regarding physical effort (eventually signallers will get tired if a signal is very long), constraints of breath (in the case of speech) and memory constraints. These latter points perhaps have less effect in language change since existing word-forms in modern day languages are so much shorter than they would need to be before we ran out of energy, breath or memory. However, it is not clear whether this was always the case, and even small effects can be amplified through cultural transmission. In any case, because compositional structure is typically constructed from bigger chunks of signal, it is perhaps more likely that these pressures and constraints may have more effect at this level than at the level of combinatorial structure.

9.2.2 Time Constraints in Experiments

The effects of duration constraints on signals haven't been systematically investigated in the artificial language literature, though some experiments in the literature impose duration limits while others do not. Some experiments have no limit on the duration which signals can be (e.g. Verhoef et al., 2014) , and others have a maximum duration signals can be (e.g. Verhoef et al., 2015), while others have made signals have a set duration that they have to be (e.g. Verhoef et al., 2016). Other experiments do have a time limit, as a matter of practicality in the experimental design, but one which is so long that participants are never likely to reach it, meaning that no adaptive pressure is likely to be attached to the time limit (e.g. in Roberts & Galantucci, 2012; Roberts et al., 2015). Table 3.1 outlines experiments that use continuous signal spaces and whether they use time limits on their signals. Further, I have included information about whether the experiments in question are social coordination experiments, or iterated learning experiments. This last column is important, because when participants are learning a set of existing signals (as they would be in iterated learning), then they will subject to an implicit pressure to make their signals a similar duration to the pre-existing signals in recall.

In the existing literature that uses continuous signals, when time limits do exist, or indeed when they are absent, it is very rarely justified why these choices have been made, or discussed how these choices could affect the results. One related area that has been explicitly discussed is rapidity of fading.

9.2.3 Rapidity of fading

As discussed in Chapter 2, rapidity of fading is very related to signal duration. Galantucci et al. (2010) is the only experimental study that has looked into rapidity of fading. In that study, signals in the slow fading condition disappeared after 2.5 seconds, but

Study	Duration limit	Structure
Galantucci (2005)	No limit stated	Social Coordination
Verhoef (2012)	No limit	Iterated Learning
Del Giudice (2012)	No limit stated	Iterated Learning
Roberts & Galantucci (2012)	$\sim$1 minute maximum	Social Coordination
Verhoef et al. (2015a)	No limit	Iterated Learning
Verhoef et al. (2015b)	4 second maximum	Social Coordination and Iterated Learning
Roberts et al. (2015)	$\sim$1 minute maximum	Social Coordination
Verhoef et al. (2016)	5 seconds exactly	Social Coordination

Table 9.1: A list of experimental studies using continuous signals and whether they have duration constraints on the the signals produced. Studies are ordered by the date of publication.

there was no reported limit on the duration signals could be. Nearly all other experimental studies looking at the emergence of combinatorial structure, especially when the signals are auditory in nature, have very fast rapidity of fading where signals disappear as soon as they are seen/heard.

9.3 Hypotheses

The two possible mechanisms outlined above for how limiting signal duration will affect structure in signals are as follows:

1. Enabling longer signals will facilitate the emergence of combinatorial structure. This is due to the ability to use the time dimension to get around crowding in a signal space.

2. Restricting signals to be shorter will cause more pressure for combinatorial structure. This is due to pressures for efficiency and information density. These pressures will reduce the amount of redundancy in signals, which may facilitate the process of analysing reusable chunks as structure.

These mechanisms are not the only things affecting the emergence of combinatorial structure and they are not actually contradictory or mutually exclusive. Hypothesis

1 may be at play during the initial emergence of combinatorial structure, but then hypothesis 2 becomes more important once a communication system has been established. Participants in the current experiment are generating systems from scratch, and so perhaps hypothesis 1 is more pertinent. In order to investigate hypothesis 2, I may need to artificially overcrowd the signal space in other ways (either with more meanings or with noise).

9.4 Methods

9.4.1 Conditions

Participants took part in one of two conditions:

1. The constrained condition, signals could only be 1 second long.

2. The unconstrained condition had no time limit on signals. Data from Experiment 5 was used for this condition.

9.4.2 Participants

In the constrained duration condition there were 18 participants, 12 female, 6 males. Average age 25.9 ($SD = 4.9$). Participants were recruited at the Vrije Universiteit Brussel and the Katholieke Universiteit Leuven. Participants were paid €5 for the 30 minutes it took to complete the experiment. I collected data on their musical proficiency (self reported on a 5 point scale from "none" to "expert"), as well as the languages they speak and their proficiency levels in each. All participants gave informed consent to have their data recorded and knew they could leave the experiment at any point.

For participant data for the unconstrained condition, see information reported in Chapter 8, Experiment 5.

9.4.3 Signals

Signals were produced in the same way they were in Experiments 5 and 6, with the horizontal axis of the Leap Motion, left to right, affecting the pitch low to high respectively. However, in the constrained condition of this experiment, signals could only be 1 second long. Participants were told explicitly that their signals could be no more than 1 second long and they were given a progress bar (see figure 9.1) so they could see the second passing.

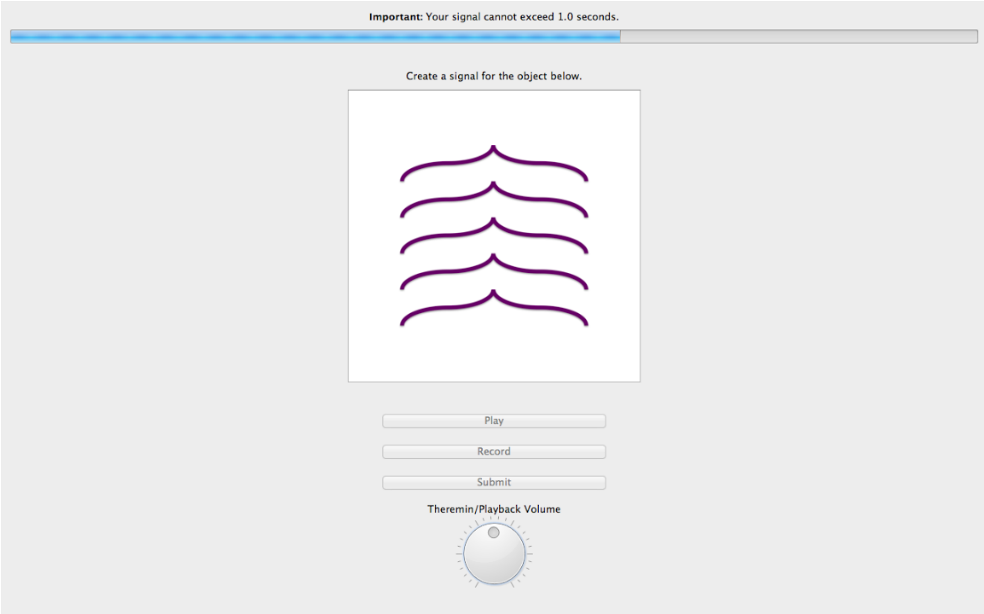

Figure 9.1: The interface during the signal creation section of the experiment. The progress bar counting down 1 second can be seen at the top. Participants were asked to rerecord if the signal was empty, or they could rerecord if they were not happy.

Participants could listen back to their signals and rerecord them if they were unhappy. This was particularly important in this experiment with only 1 second of time available to get it right.

One second is a very short amount of time. However, the aim of this experiment is to identify the effects of duration constraints. In Experiment 5, despite there being no limit on duration, signals remained quite short, with a median length of 2.2s (Q1(1.1s), Q3(3.6s)). In the current experiment, I halved this median length in order to create a real time pressure.

9.4.4 Meanings

The meanings in this experiment were the same as the meanings used in the experiments in chapter 8.

9.4.5 Procedure

In order to make the data comparable, this experiment had exactly the same procedure, structure and phases as Experiment 5. However, only data produced in the last phase of Experiment 5 and this experiment were used in the analysis, unless specified otherwise.

9.4.6 Post-experimental Questionnaire

As in the other signal creation experiments in this thesis, I gave all participants a post-experimental questionnaire asking whether they found the experiment easy and what strategies they used to produce and recognise their signals in each phase of the experiment.

9.5 Results

9.5.1 Qualitative Analysis

Before going into a more quantitative analysis, this section briefly gives an example of the types of signals and repertoires produced in each condition.

A full signal repertoire from both conditions can be seen in figure 9.2. The effects of signal constraints are very clear in the differences between these two signal sets. In the constrained condition (above) signals, especially in the first phase, participants rely on stable pitches. As the experiment continues, more movement is introduced into the signals. In contrast, in the unconstrained condition (below) signals display much more movement and a larger variation of strategies to create distinctions.

9.5.2 Quantitative Analysis

I used a linear model with musical proficiency as the independent variable and performance (score) in the signal recognition task as the dependent variable, and found that there was no predictive relationship ($F(1,16) = 2.1, p = 0.16$).

In the constrained condition, participants were worse at recognising their own signals ($M = 64\%$ correct), than in the unconstrained condition ($M = 86\%$ correct). This discrepancy in success indicates that in the constrained condition, the participants had a much harder time creating distinct signals. In the constrained condition, signals were

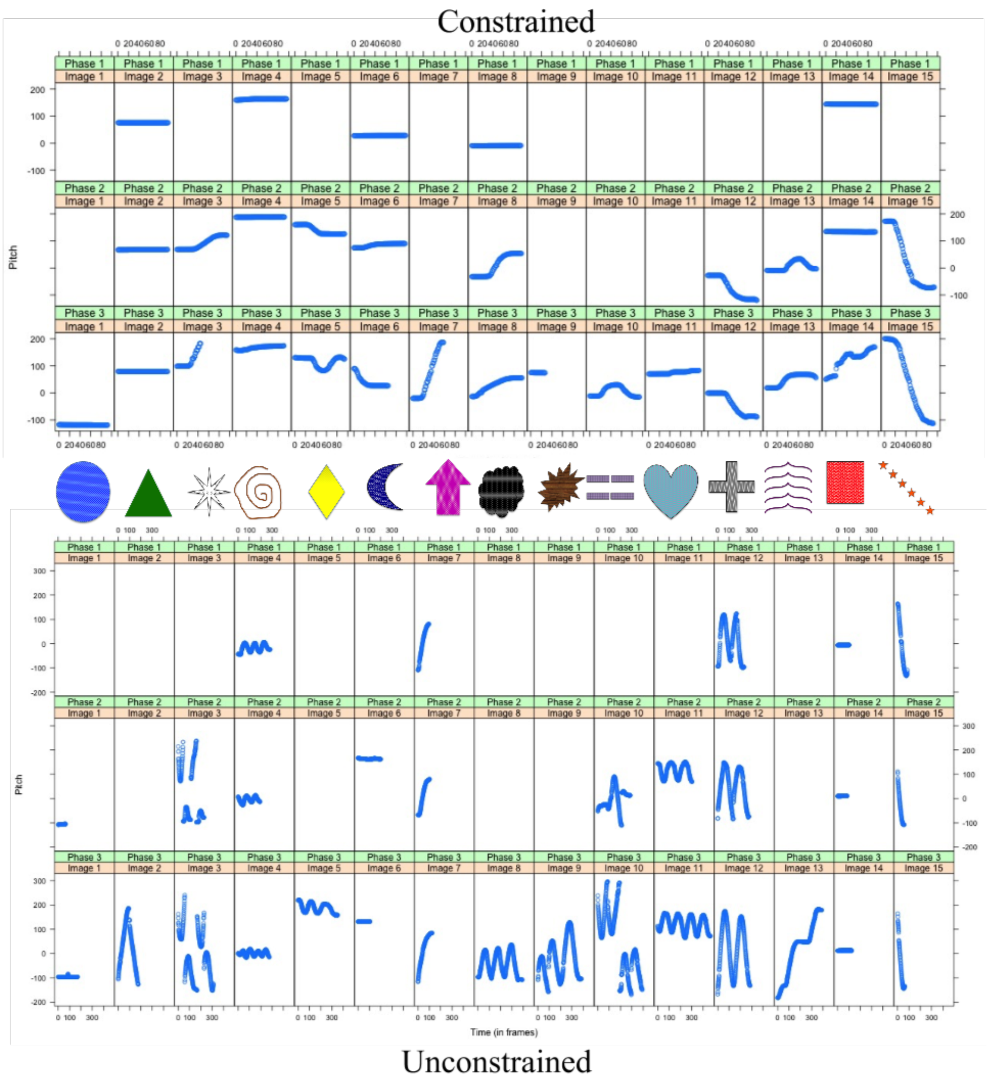

Figure 9.2: Examples of signal repertoires produced by one participant in the constrained condition (above) and one participant in the unconstrained condition (below). Each row of signals is the last production from each phase. The meaning that each signal refers to lines up with the images in the centre of the figure. The y axis represents hand position in coordinates. The x-axis is time (in data frames, 110df/s). However, the signals in the constrained repertoire are plotted with each x-axis being one second. The signals in the unconstrained repertoire are plotted with each x-axis being 5 seconds (the maximum length of signal in this particular repertoire). In reality, the duration of signals in the unconstrained condition were much longer.

much simpler, with a lot of participants relying on static pitch, rather than on patterns and pitch changes. As in previous chapters, I measured the amount of movement within signals by calculating the variance of the signal trajectory coordinate values. I also built a mixed effects model with phase as a fixed effect and participant number and meaning as random effects. With condition as a predictor variable, and comparing with a null model without condition, I found that standard deviation in signal trajectories was significantly lower in the constrained condition ($M = 0.2s$), than in the unconstrained condition ($M = 0.4s$) ($\chi^2(1) = 9$, $p < 0.001$). I also measured predictability in signals, controlling for the same effects as above, and again comparing with a null model without condition. I found that condition did predict the amount of predictability in signals ($\chi^2(1) = 26.5$, $p < 0.001$), with signals in the constrained condition being more predictable. However, this effect went away completely when I controlled for duration of signals.

As reported in Chapter 8, there was a significant trend in the amount of predictability in signals decreasing (so being more complex) as the meaning space expanded in the unconstrained condition. This trend did not occur in the constrained condition, suggesting that the limited signal duration stopped participants from creating new strategies for new meanings, or possibly constrained the emergence of redundant features in new signals.

9.5.3 Post-experimental Questionnaire

As in previous experiments, participants attempted to employ iconic strategies. One particularly imaginative participant wrote “for the heart, it was a gradual decrease of pitch, considering a heartbreak. For stars and clouds, high pitch as they are far away from us.” However, in the constrained condition, participants were much more likely to write that they relied on the pitch of signals to define them, rather than the shape or pattern of their hand movement.

9.6 Discussion and Conclusion

In this chapter I have attempted to demonstrate the effects that duration constraints will have on signal structure. Though the experiment was very simple, it did present some interesting results. First, signals produced in the unconstrained condition were a lot more complex than signals in the constrained condition. This is perhaps simply because they had more space on the time dimension to be more complex, which also acted as a way to aid discrimination between signals. In the constrained condition even if participants moved their hand more quickly to cram more signal elements in, there was a limit to what they could do in the allotted time. This hypothesis is supported by the fact that the signals grew in complexity as the experiment progressed and the meaning space expanded only in the unconstrained condition, suggesting that this was happening in order to maintain discrimination.

These results lend some support to hypothesis 1, that longer signals allow for the use of the time dimension to aid discrimination and get around the crowding in a signal space. In the current experiment, this is not necessarily linked to the emergence of combinatorial structure as I have no results explicitly measuring the reuse of building blocks. However, I would expect more complex signals to be a better starting point for combinatorial structure to emerge as it gives more potential for reanalysis and reuse of building blocks.

In the constrained condition, where signals were largely constrained to encoding meaning with pitch or very simple forms of movement within the signals (pitches moving from up to down or vice versa, for instance), there was very little opportunity for concatenation of signals because there wasn't the time to do it. However, restricting the time dimension would perhaps cause more pressure to categorise the signal space more quickly as this was the primary way to create signal distinctions. A more targeted experimental design would perhaps be able to get at whether time constraints facilitate the categorisation of a signal space, as the current experiment only provides a first step to understanding what effect time constraints can have. It would also be insightful to perhaps run a version of this experiment where the signal space is artificially crowded by noise to see if this drives signals to be longer.

The results here encourage theoretical consideration of the links between signal duration and combinatorial structure. Perhaps more importantly in the current literature, though, they highlight why experimental studies need to consider the effects that time constraints will have on structure. While no previous experimental work has limited signals to such an extremely short duration as one second, having short limits, or specific durations that signals have to be, has the potential to influence the structure that can emerge in those signals.

CHAPTER 10

Conclusion

The experiments and theory covered in this thesis broadly address the effect of modality on how structure in language emerges. The conclusions of this work are important in 2 main areas. The first is research into the emergence of structure in the language of our ancestors, and also how structure emerges and evolves in existing real world languages. Second, the thesis has significant implications for the design of artificial language experiments.

This thesis opened with the observation that languages are made up of a finite number of meaningless building blocks, termed throughout as *combinatorial structure*. This level of structure is almost ubiquitous in the world's languages. However, there are some examples of languages which exist, or have existed, without it. These languages are emerging signed languages in rural signing communities. Even in fully developed sign languages, combinatorial structure is less prevalent, with fewer minimal pairs and less internal structure within lexical items, than in spoken language (Sandler, 1996). As a result, researchers usually account for the differences in prevalence of combinatorial structure as being the result of the modality used (speech or sign). However, precisely what it is about the modality that causes the differences is often disputed. In this thesis, I have systematically investigated various reasons for why different language modalities might cause the differences we see in the emergence of combinatorial structure in the real world.

Further to real world implications (summarised in Chapter 3), there is a vast range of proxies used for signalling spaces in artificial language experiments, summarised in

Chapter 4. These proxies all have different physical properties and act as models for language modalities in the real world. The use of experimental proxies is necessary to stop interference from existing linguistic knowledge. However, this then creates issues of ecological validity. The more similar a proxy is to a signalling modality, the more interference you will get. However, the more different a proxy is to a real world modality, then the less relevant your result may be to real-world linguistic behaviour. As with modalities in the real world, artificial proxies used in experiments also come in varying shapes, sizes and modalities (visual, auditory, tactile, etc.) and have different degrees of mappability and noise. Accordingly, it is necessary to understand how the physical differences between artificial proxies affects the structure in artificial signals. Understanding this will then make it easier to identify effects that exist on top of physical effects, facilitating the ability to distinguish whether the emergence of structure is the result of cognitive or cultural biases.

10.1 Findings

Here, I briefly summarise the findings from each chapter.

First, in Chapter 5, I investigated the overall size of the meaning space. It is difficult to quantify size in either real world languages or in a controlled experiment. The size of the production space does not represent the size of the perceptual space. Furthermore, it is difficult to measure the size of a perceptual space. I attempted to restrict only one thing about a production space (the pitch range of a slide whistle), in order to affect the size of the perceptual space. Despite the controlled nature of this experiment, it is still unclear quite how this manipulation affected the size of the perceptual signalling space. Restricting the pitch range changed the amount of static pitches that could be perceptually differentiated, but did not change people's ability to replicate pitch patterns. Whether the size of the signalling space was restricted or not, then, depends on what people were using to differentiate signals: absolute pitch or patterns in pitch. In this task, it seemed to be the latter. This resulted in participants behaving in very similar ways in both the restricted condition and unrestricted condition.

In Chapter 6, I explored the dimensionality of a signalling space, or the number of ways that differences can be made between signals, directly addressing the problem with the experiment in Chapter 5. That is, instead of manipulating the physical size of a signalling space, I manipulated the number of ways that signals could be differentiated. I found that participants use iconic strategies when the number of signal dimensions matches the number of meaning dimensions. There was more evidence of movement and complexity in signals when the meaning dimensions outnumbered the available signal dimensions. In Chapter 3, I discussed evidence that the manual modality has more signal dimensions than the spoken modality. I argue that the dimensionality offered by a modality may be at the root of why combinatorial structure emerges more slowly in the manual modality. I also outline that the number of dimensions available in a signal space proxy will affect experimental results, especially in relation to the number of dimensions a meaning space has. Both signal spaces and meaning spaces are heavily controlled in experiment design. The results I present here outline why these dimensionalities, and how they might map to one another, are very important to consider, both in the design and interpretation of results from experiments. However, more work may need to be done to work out whether dimensionality in both signal spaces and meaning spaces is better thought about as a physical property of the world, or more in terms of how individuals conceptualise the dimensionality of signal spaces and the world at a cognitive level. These questions don't necessarily negate the findings of the experiments in Chapter 6, as the signal space and meaning spaces in the experiments were very small and controlled, making it less likely that there was a difference between their physical properties and how they were cognitively perceived. Further, I do not think humans are born with conceptual dimensions for either signal or meaning spaces, and therefore there is potentially less of a confound from them being different in experiments that are looking at systems being created from scratch.

In Chapter 7, I looked at different ways that signal-meaning mappings can be disrupted and how different types of iconicity might affect signal structure in different ways. I found that when a continuous signal space is used to describe a continuous meaning space, then participants adopt signals with diagrammatic iconicity. However, when participants are to create signals for a discretised meaning space, they adopt more abstract forms of iconicity. These more abstract forms have more complexity than the diagrammatic forms. This raises some interesting implications on a few levels. The first is that different signalling spaces might inhibit complex structure in signals if they afford more diagrammatic iconicity (but not necessarily imagic iconicity). The second is that different meaning spaces may facilitate or inhibit conventionalisation and the emergence of combinatorial structure. This experiment also showed that participants were better at differentiating signals that did not use diagrammatic iconicity, but that mistakes within systems with diagrammatic iconicity were relatively small and predictable. This hints that while diagrammatic iconicity is useful for productivity and transparency in signals, it may not be a good strategy for accuracy of expression.

In Chapter 8, I explored conventionalisation explicitly in the context of comparing a communication task with an individual signal creation task. In the literature there is an oft cited finding that repeated interactions cause signals to lose iconicity (Garrod et al., 2007). However, I found the opposite of this: in the communication condition of my experiment, repeated interactions caused iconicity to go up rather than down, which did not happen in the individual condition. I link this finding directly to the modality used. Modalities that facilitate iconicity will allow for signals that are highly iconic as a starting point, which enables transparency for all communicators from the beginning. However, when signals cannot start as transparent because the modality does not lend itself well to iconicity, then interaction acts as a way for signals to grow in iconicity, rather than lose iconicity. However, there was another aspect of this experiment that may have contributed to the results seen, which was that the meaning space was growing as the experiment went on. This expanding meaning space may have been driving signals to grow in complexity in order to be differentiated, however this does not explain the difference between the conditions, as both conditions had an increasing meaning space. Given this, regardless of the underlying mechanism, the findings are likely the result of the modality being used, as the restrictive pitch signals do not lend themselves to complexity as a starting point.

Finally, in Chapter 9, I presented a brief experimental study looking at the effects of duration constraints on signals. This was primarily a study testing the assumption that having a time limit on signals will not affect the structure that comes out of them. Enabling longer signals will facilitate the emergence of combinatorial structure. I tested whether restricting duration of signals would increase the need for efficiency and the density of information in signals and decreases the presence of redundancy in signals, which would facilitate the emergence of combinatorial structure. On the other hand, allowing longer signals may be exactly the thing that facilitates combinatorial structure because it allows for compensating for crowding in the signal space along the time dimension. The results showed that signals produced without a time constraint had more complexity than signals produced with a time constraint, and that people were better at differentiating the complex signals in the unconstrained condition. This indicates that extra time was being used to makes distinctions clearer between signals, lending support to the hypothesis that extra time was facilitating something like combinatorial structure.

The investigations throughout this thesis have been semi-successful in identifying mechanisms that may have facilitated or affected the emergence of combinatorial structure, though how to characterise the structure that have emerged in the experiments has not always been clean-cut and often confounded with things like iconicity. The experiments in the thesis only show the very early stages of linguistic grounding, and combinatorial structure is not, I think, something that comes as a starting point. Combinatorial structure emerges as a result of holistic signals being reanalysed into smaller building blocks, or from existing compositional building blocks losing their meaning. While I have not been able to explicitly demonstrate such processes (as a result of my experiments starting from scratch and being relatively short), I feel I have shown some conditions where the emergence of combinatorial structure may be more likely which I have identified when signals have been longer or more systematic or display more movement.

10.2 Existing Questions and Further Work

This thesis has by no means been a comprehensive work on all of the ways in which modality may affect structure, nor a comprehensive investigation of the modality effects it has identified. Below is a collection of possible (though by no means comprehensive) avenues for further research that directly stems from the work presented in this thesis.

10.2.1 Iconicity

A lot of current experimental work, including the work presented in this thesis, is concerned with the idea that combinatorial structure comes from not being able to rely on iconic strategies. This creates an assumed dichotomy between iconicity and combinatoriality in signals. However, it is not true that a signal cannot be both iconic and have combinatorial structure. The relationship between these two phenomena and how they influence each other, then, is something that needs further investigation both from theoretical and empirical perspectives. Does combinatorial structure make signals less likely to be interpreted as iconic? Does iconicity inhibit the emergence of combinatorial structure? Can combinatorial structure and iconicity reliably co-exist in a system of signals? While this thesis, and existing literature, has touched on some of these questions, the discussion around these issues relies on a number of assumptions that require proper inspection. I am interested in pursuing these issues in relation to types of iconicity present in a system (i.e. imagic vs. diagrammatic), iconicity in relation to the crowding of signal (or semantic) spaces, and whether iconicity exists at the level of production and/or perception. I will detail some possible research avenues relating to all of these ideas below.

10.2.1.1 Types of Iconicity

I spent some time exploring different types of iconicity in Chapter 7, and showed how different forms of iconicity may provoke very different types of structure and iconicity in signals. However, this experiment was only demonstrating the behaviour of an individual creating signals from scratch. I would be very interested to see this work be progressed using communication games or iterated learning to see how different types of iconicity conventionalise. The hypothesis I put forward in Chapter 7 was that more abstract forms of imagic iconicity would lend themselves to conventionalisation a lot easier than diagrammatic iconicity. I would be interested to investigate this experimentally.

10.2.1.2 Crowded Semantic Spaces

In the first phase of Experiment 2, the squares differed in 5 gradations of size, but in the other two phases, the squares only differed along 3 gradations of size and shade. This had the potential to contribute to effects of semantic space crowding. In chapter 2, I discussed the hypothesis of Gasser et al. (2010), that when a semantic space gets crowded, then this will generate a stronger pressure for expressivity than is normally present, because if iconic signals are very similar to one another because their meanings are very similar to one another, then they will be more easily confused. Of course, this may not be problematic if the meanings are similar enough, and may simply cause semantic categories to form. However, with the correct communicative pressure, signals may become distinct in ways that are not iconic. I do not think that the distinction between 3 and 5 squares in Experiment 2 caused this effect of semantic space crowding, indeed the analysis showed that signals displayed diagrammatic iconicity in both phases. However, I am interested in exploring this effect of semantic space crowding using more dramatic comparisons, perhaps using and experiment very similar to Experiment 3. Such an experiment could inform computational work modelling how confusability between meanings/signals might scale up and affect systematicity at bigger scales. Further, I would like to see the effects of changing pressures on communication, e.g. affecting whether meanings need to be identified from signals in an accurate manner, or just be “good enough”.

10.2.1.3 Iconicity in Production vs. Perception

In Experiment 7, I measured the iconicity of the audio signals produced by participants in both Experiments 5 and 6. However, as discussed in Chapter 3, the Leap Motion paradigm has the property of having a strong distinction between the production space (the physical space where a participant moves their hand), and the perception space (the auditory space that is used to transmit the signal). As a result of this, if someone attempts to draw a shape with the paradigm (such as one used in Experiment 5 or 6), then they are possibly more likely to use the visual articulation space than trying to produce an iconic signal that is apparent from the auditory signal. This is possibly more true at the beginning of the experiment before they have become accustomed to the fact that the information transmitted is the auditory signal, rather than the gesture that produced it. There is some evidence from the post-experimental questionnaires that this is the case, with some participants stating that they started by trying to “draw” the shapes, and later adopted other strategies that relied more on audio cues, such as making a certain number of movements that corresponded to the number of corners in a shape. This trend would explain the rise in iconicity in signals produced later in Experiment 6, when compared with signals produced earlier. If participants playing the communication game changed their strategies to be more iconic from an auditory perspective, then naïve participants, who are only exposed to the auditory signals, would only be susceptible to the auditory iconicity, and not the visual iconicity that the participants potentially started with. In order to test this hypothesis, a further playback experiment is being conducted (initial results are available in Little et al., 2017c), where naïve participants are exposed to the visual representation of the hand movements of participants instead of the audio signals produced. It is possible that naïve participants’ ability to match these visual signals to their intended referents will be better for signals produced at the beginning of the experiment.

10.2.2 The Effect of Communication

In this thesis, I have presented several individual signal creation experiments (Experiments 2, 3, 4, 5 and 8). These experiments had a pressure for expressivity where participants “communicated” with themselves by recognising their own signals. While this mechanism does create a pressure to keep signals distinct, there are several reasons why having a communication partner within such experiments might make a difference to the signals produced. Some of these reasons were discussed in Chapter 8, as that chapter dealt with a comparison between an individual signal creation experiment (Experiment 5) and a communication game using the same set of meanings (Experiment 6). However, I'd like to discuss here why communication might affect the signals that emerged from experiments 2, 3 and 4 specifically. These experiments all dealt with iconicity of different types, and post-experimental questionnaires illustrated that participants frequently perceived the signals they were producing as iconic. Despite this, the diversity in the strategies used was notable, which caused a problem for analysis as the more diverse a behaviour is, the more difficult it is to model it using statistics. However, a language is limited in how diverse it can be in order to be intelligible by all of the people who speak it. That is, the knowledge that is used to create signals that are to be understood by another person will be restricted by what knowledge is shared, rather than relying on individual, idiosyncratic intuitions. This should reduce the diversity in iconic signals because they will be making use of information from a smaller pool of shared knowledge. As such, I think running these experiments as communication games with more than one participant will produce very different results, that are perhaps more ecologically valid.

Another way that introducing communication to experiments might affect the signals comes from an artefact of the Leap Motion paradigm. The signals produced using the Leap Motion are both visual and auditory. The participant can hear the signal produced by their hand movements, but they can also see their hand in front of them as they produce it. When they "recognise" their signals, in both the individual tasks, and in a communication game, they only receive the auditory information. In the individual task, they have a memory of all of the hand movements they made to produce the signals, which will make the task of recognising their signals easier. However, in a communication game, the task of recognising an auditory signal based on hand movements that have not been seen by the other participant will be more difficult. Further, if participants want to use the same signal as their partner, the communication game calls for participants to reconstruct auditory signals into hand movements based on auditory information alone. This reconstruction is not necessary in the individual experiments as the participant has the memory of the hand movements they used to produce the signals in the first place. More work needs to be done in order to make a comparison between individual conditions and communication conditions, perhaps by including a third condition where participants can see each other.

10.2.3 The Effect of Transmission

Much experimental work in evolutionary linguistics has demonstrated that for compositional structure to emerge, one needs both pressures for expressivity from communication and pressures for learnability from transmission (Kirby et al., 2015). This thesis included one iterated learning study (in Chapter 5) and one communication game (in Chapter 8) but a comparison was not made between these two paradigms. This comparison is not possible in the current thesis as the two experiments were too different from one another (different signalling spaces, one had a meaning space and the other did not, different numbers of signals were produced, etc.). While other experiments have used slide whistle signals in iterated learning experiments (e.g. Verhoef et al., 2014) and some have used them in communication games (Fullerton, 2011), a comparison has never been made between these two paradigms in the same study, and so this is an obvious avenue for future research.

10.2.4 Measuring how many dimensions there are in speech or the manual modality

I discussed in Chapter 3 my argument that the manual modality has more dimensions than the spoken modality. However, this remains quite a controversial claim. As well as considerations for the number of dimensions possible, there are also considerations to be made about the number of degrees of freedom within those dimensions. To my knowledge, no study has ever tried to quantify the magnitude of possible distinctions possible within both the spoken and manual modalities. Such a study would be a necessary addition to the available evidence regarding claims of how exactly to quantify the size differences between the modalities.

10.2.5 Noise

As discussed in Chapter 2, having noise in a communication channel is likely to promote the emergence of combinatorial structure because it will affect how quickly a signal space becomes crowded. The effect of noise on the emergence of combinatorial structure has been explored using computational modelling (Nowak et al., 1999; Tria et al., 2012), however, it has not been investigated in experimental work. The effects of noise could be investigated using the Leap Motion paradigm by designing a communication game where noise is added, or signals are altered in some way, as they are transmitted from one participant to another.

10.2.6 Experimentally Modelling the Emergence of Combinatorial Structure from Compositional Structure

One of the main hypotheses in this thesis is that compositional structure likely predates combinatorial structure, given the right modality capabilities. While various experiments throughout the thesis have tested the plausibility of the hypothesis that combinatoriality emerged from a compositional or iconic system, much of the structure seen in the signals produced remains ambiguous. Throughout this thesis, it has been difficult to draw lines between structure that is combinatorial, compositional, or in some instances, iconic. Further, much, if not all, existing experimental work focuses on the emergence of structure on only one level at a time. I would like to briefly outline an experiment that aims to model the emergence of structure on both levels, while keeping the levels of structure definable and distinct.

The experiment would be a communication game where the meaning space expands in a similar way to how the meaning space expanded in Experiment 6. However, in Experiment 6, the meanings were designed to have no internal structure and no shared features. If, instead the meaning space started as being very small and structured (for example, one or two shapes with one or two different colours) then a very compositional system will be the easiest strategy for participants, with perhaps one signal unit that refers to colour and one that refers to shape. The experiment would then gradually add more and more meaning dimensions until a compositional system is no longer feasible, and units need to be combined in potentially more combinatorial-like ways.

10.2.7 Computationally Modelling the Emergence of Combinatorial Structure from Compositional Structure

Some computational agent-based models exist that explore the emergence of duality of patterning (e.g. Tria et al., 2012), but modelling work so far has focused on one variable (e.g. noise, dimensionality, communication, iteration etc.). In many respects, computational modelling is a lot better suited than other methods to answering big question problems of how structure emerges at 2 levels, because of its ability to run big populations and many generations quickly. I would like to see further work using data from real world languages (especially emerging sign languages) and the results from the experiments I have presented in this thesis (as well as experimental work by others) to inform parameters in agent-based models. We need to consider many variables to fully understand how and why structure emerged in language, and gradually more and more complex models can help us to move towards a state of full understanding. I think the set of variables discussed throughout this thesis all have their place in the bigger picture and I would like to see them embedded in bigger and more ambitious work.

10.3 Overall Conclusion

Taken together, the big picture to which the experimental results presented in this thesis contribute is that of the many ways in which modality might affect the emergence of structure in signals. These effects come in two main categories: those concerning signal space crowding (e.g. effects of number of meanings, signal space size and dimensionality, noise, etc.) and those concerning signal-meaning mapping (matches in modality, dimensionality and mappability) - though these two categories have significant overlap and cannot be easily teased apart. Through controlled manipulation of one main paradigm, I have added to the existing evidence about the effects of transmission (championed by Verhoef et al., 2014) and communication by considering the effects of the modality used. I would not like to claim that my findings and hypotheses are in opposition to narratives stressing the importance of transmission and use, but instead I would argue that I am adding to the rich tapestry of things that need to be considered in order to get a full picture of how language emerged. I hope I have made a convincing case that the emergence of structure is not something that is the result of one evolutionary pressure, but something that is affected by functional and cognitive pressures, as well as modality effects.

Having stressed that modality effects are merely one part of a bigger picture, I do think they are more important regarding some evolutionary questions than others. Specifically, the question of whether combinatorial or compositional structure came first in language. I argue that modality effects are central to this question as the number of holistic signals a modality can communicate directly affects the order of emergence. If a modality allows for many thousands of holistic signals (as it is possible that the manual modality does), then a system is likely to adopt compositional structure first, as in emerging signed languages. If, on the other hand, a modality is a lot more restrictive in the amount of distinct signals it can produce, then combinatorial structure will become necessary a lot more quickly in order to aid this discrimination. It may be important then, when looking at the abilities of animals and structure in animal communication systems, to consider the flexibility of the modality the animal is using, before accounting for structure at the level of cognition.

The specific contribution of this thesis to the existing literature on language evolution has been consideration of how modality may affect the emergence of structure in real world languages, as well as in lab-based experiments. One recurring issue in the current literature on evolutionary linguistics is a tendency for a finding to be cited and built upon without replication and, more often than not, it is assumed that a finding in one modality is likely to hold for another (in both real world data and lab experiments). As a relatively new field (especially regarding the use of continuous signal spaces) there are very few experiments that have been carried out with multiple signalling paradigms, making it difficult to know the effects of a particular signalling space or modality. This thesis has attempted to demonstrate that the signalling space does impact structure and iconicity in signals, and has outlined the need for replication using different modalities, and consideration for modality effects in experimental design.

References

Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures*, *12*(1), 1–13.

Arnold, K., & Zuberbühler, K. (2006). The alarm-calling system of adult male putty-nosed monkeys, cercopithecus nictitans martini. *Animal behaviour*, *72*(3), 643–653.

Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: compositional or idiomatic expressions? *Brain and language*, *120*(3), 303–309.

Bartlett, F. C. (1933). Remembering: A study in experimental and social psychology. *British Journal of Educational Psychology*, *3*(2), 187–192.

Battison, R. (1978). *Lexical Borrowing in American Sign Language*. ERIC.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, (p. 201605782).

Bloomfield, L. (1933). *Language*. Henry Holt.

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.

Brentari, D. (2002). Modality differences in sign language phonology and morphophonemics. *Modality and structure in signed and spoken languages*, (pp. 35–64).

Brighton, H., Kirby, S., & Smith, K. (2005). Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable. *Language origins: Perspectives on evolution*, (pp. 291–309).

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*.

Caselli, N., Ergin, R., Jackendoff, R., & Cohen-Goldberg, A. (2014). The emergence of phonological structure in central taurus sign language. In *From Sound to Gesture*. Padua, Italy.

Chomsky, N. (2016). The language capacity: architecture and evolution. *Psychonomic Bulletin & Review*, (pp. 1–4).

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and brain sciences*, *31*(05), 489–509.

Collier, K., Bickel, B., van Schaik, C. P., Manser, M. B., & Townsend, S. W. (2014). Language evolution: syntax before phonology? *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1788), 20140263.

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 35th Annual Conference of the Cognitive Science Society*, (pp. 340–345). Austin, Tx: Cognitive Science Society.

Crasborn, O., van der Hulst, H., & van de Kooij, E. (2002). Phonetic and phonological distinctions in sign languages. In *A paper presented at Intersign: Workshop*, vol. 2.

de Boer, B. (2016). Evolution of speech and evolution of language. *Psychonomic Bulletin & Review*, (pp. 1–5).

de Boer, B., Sandler, W., & Kirby, S. (2012). New perspectives on duality of patterning: Introduction to the special issue. *Language and cognition*, *4*(04), 251–259.

de Boer, B., & Verhoef, T. (2012). Language dynamics in structured form and meaning spaces. *Advances in Complex Systems*, *15*(3), 1150021–1–1150021–20.

De Saussure, F. (1983). *Course in general linguistics*. Columbia University Press.

Del Giudice, A. (2012). The emergence of duality of patterning through iterated learning: Precursors to phonology in a visual lexicon. *Language and cognition*, *4*(4), 381–418.

Del Giudice, A. (2012). The emergence of duality of patterning through iterated learning: Precursors to phonology in a visual lexicon. *Language and cognition*, *4*(4), 381–418.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, *6*(10), 654–672.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences*, *19*(10), 603–615.

Emmorey, K. (2002). The effects of modality on spatial language: How signers and speakers talk about space. In R. P. Meirer (Ed.) *Modality and structure in signed and spoken languages*, (pp. 405–422). Cambridge, Cambridge University Press.

Eryılmaz, K., & Little, H. (2016). Using leap motion to investigate the emergence of structure in speech and language. *Behavioral Research Methods*, (pp. doi:10.3758/s13428–016–0818–x).

Eryılmaz, K., Little, H., & de Boer, B. (2016). Using hmms to attribute structure to artificial languages. In S. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher, & T. Verhoef (Eds.) *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*, (pp. 407–409).

Esper, E. A. (1925). A technique for the experiment investigation of associative interference in artificial linguistic material. *Language monographs*, *1*, 1–47.

Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive science*, *37*(7), 1356–1367.

Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3553–3561.

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, *34*(3), 351–386.

Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014a). Creating a communication system from scratch: Gesture beats vocalization hands down. *Frontiers in Psychology*, *5*(354).

Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014b). Creating a communication system from scratch: gesture beats vocalization hands down. *Frontiers in Psychology*, *5*, 354.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in cognitive sciences*, *4*(7), 258–267.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*(1), 173–215.

Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press.

Frishberg, N. (1975). Arbitrariness and iconicity: historical change in american sign language. *Language*, (pp. 696–719).

Fullerton, L. (2011). *The evolution of combinatorial phonology in slide whistle communication*. Master's thesis, The University of Edinburgh.

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.

Galantucci, B., Garrod, S., & Roberts, G. (2012). Experimental semiotics. *Language and Linguistics Compass*, *6*(8), 477–493.

Galantucci, B., Kroos, C., & Rhodes, T. (2010). The effects of rapidity of fading on communication systems. *Interaction Studies*, *11*(1), 100–111.

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.

Gasser, M., Sethuraman, N., & Hockema, S. (2010). Iconicity in expressives: An empirical investigation. In S. Rice, & J. Newman (Eds.) *Experimental and Empirical Methods in the Study of Conceptual Structure, Discourse, and Language*, (p. 163–180). CSLI Publications. Stanford, CA.

Goldin-Meadow, S., & McNeill, D. (1999). The role of gesture and mimetic representation in making language the province of speech. In M. C. Corballis, & S. Lea (Eds.) *Evolution of the Hominid Mind*, (pp. 155–172). Oxford University Press.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive Science*, *31*(3), 441–480.

Haiman, J. (1985). *Iconicity in syntax: proceedings of a Symposium on iconicity in syntax, Stanford, June 24-6, 1983*, vol. 6. John Benjamins Publishing.

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., Chomsky, N., & Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in psychology*, *5*, 401.

Healey, P. G., Swoboda, N., Umata, I., & Katagiri, Y. (2002). Graphical representation in graphical dialogue. *International Journal of Human-Computer Studies*, *57*(4), 375–395.

Healey, P. G., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. *Cognitive Science*, *31*(2), 285–309.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*, 88–111.

Holz, D. (2014). Systems and methods for capturing motion in three-dimensional space. US Patent 8,638,989.
URL `https://www.google.com/patents/US8638989`

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1651), 20130298.

Janssen, R., Winter, B., Dediu, D., Moisik, S., & Roberts, S. (2016). Nonlinear biases in articulation constrain the design space of language. In S. G. Roberts,

C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher, & T. Verhoef (Eds.) *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*, (pp. 448–451).

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, *165*, 45–52.

Kirby, S. (2002). Natural language from artificial life. *Artificial life*, *8*(2), 185–215.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Kirby, S., Smith, K., & Brighton, H. (2004). From ug to universals: Linguistic adaptation through iterated learning. *Studies in Language*, *28*(3), 587–607.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Kirchner, R. (1998). *An effort based approach to consonant lenition*. Ph.D. thesis, UCLA.

Klima, E. S., & Bellugi, U. (1980). *The signs of language*. Harvard University Press.

Lewis, M. L., & Frank, M. C. (2016). The length of words reflects their conceptual complexity. *Cognition*, *153*, 182–195.

Little, H., & de Boer, B. (2014). The effect of size of articulation space on the emergence of combinatorial structure. In A. Cartmill, Erica, S. Roberts, H. Lyn, & H. Cornish (Eds.) *The Evolution of Language: Proceedings of the 10th international conference (EvoLangX)*, vol. 10, (pp. 479–481). World Scientific.

Little, H., Eryılmaz, K., & de Boer, B. (2015). Linguistic modality affects the creation of structure and iconicity in signals. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. Jennings, & P. Maglio (Eds.)

The 37th annual meeting of the Cognitive Science Society (CogSci 2015), (pp. 1392–1398). Austin, TX: Cognitive Science Society.

Little, H., Eryılmaz, K., & de Boer, B. (2016). Emergence of signal structure: Effects of duration constraints. In S. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher, & T. Verhoef (Eds.) *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*, (pp. 468–470).

Little, H., Eryılmaz, K., & de Boer, B. (2017a). Conventionalisation and discrimination as competing pressures on continuous speech-like signals. *Interaction studies*, *18*(3), 355 —- 378.

Little, H., Eryılmaz, K., & de Boer, B. (2017b). Signal dimensionality and the emergence of combinatorial structure. *Cognition*, *168*, 1–15.

Little, H., Perlman, M., & Eryilmaz, K. (2017c). Repeated interactions can lead to more iconic signals. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.) *the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)*, (pp. 760–766).

Little, H., Rasilo, H., van der Ham, S., & Eryılmaz, K. (2017d). Empirical approaches for investigating the origins of structure in speech. *Interaction studies*, *18*(3), 332 —- 354.

MacNeilage, P. F., & Davis, B. L. (1993). Motor explanations of babbling and early speech patterns. In *Developmental neurocognition: Speech and face processing in the first year of life*, (pp. 341–352). Springer.

Maddieson, I., & Disner, S. F. (1984). *Patterns of sounds*. Cambridge university press.

Marler, P. (1998). Animal communication and human language. *The origin and diversification of language*, (pp. 1–19).

Meier, R. P. (2002). 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 languages*, (pp. 1–25).

Meier, R. P., Cormier, K., & Quinto-Pozos, D. (2002). *Modality and structure in signed and spoken languages*. Cambridge University Press.

Meir, I., Aronoff, M., Sandler, W., & Padden, C. (2010). Sign languages and compounding. *Compounding. John Benjamins*, (pp. 573–595).

Moisik, S. R., & Dediu, D. (2017). Anatomical biasing and clicks: Evidence from biomechanical modeling. *Journal of Language Evolution*, *2*(1), 37–51.

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1152.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language. *Philosophical Transactions of the Royal Society B*.

Morgan, H. (2016). The effect of modality on signal space in natural languages. In S. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher, & T. Verhoef (Eds.) *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*, (pp. 491–492).

Motamedi, Y., Schouwstra, M., Smith, K., & Kirby, S. (2016). Linguistic structure emerges in the cultural evolution of artificial sign languages. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher, & T. Verhoef (Eds.) *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*, (pp. 493–495).

Namboodiripad, S., Lenzen, D., Lepic, R., & Verhoef, T. (2016). Measuring conventionalization in the manual modality. *Journal of Language Evolution*, (p. lzw005).

Napoli, D. J., Sanders, N., & Wright, R. (2014). On the linguistic effects of articulatory ease, with a focus on sign languages. *Language*, *90*(2), 424–456.

Nettle, D. (1995). Segmental inventory size, word length, and communicative efficiency. *Linguistics*, *33*(2), 359–367.

Nettle, D. (1998). Coevolution of phonology and the lexicon in twelve languages of west africa. *Journal of Quantitative Linguistics*, *5*(3), 240–245.

Newport, E. L., & Meier, R. P. (1985). *The acquisition of American Sign Language.*. Lawrence Erlbaum Associates, Inc.

Nowak, M. A., Krakauer, D. C., & Dress, A. (1999). An error limit for the evolution of language. *Proceedings of the Royal Society of London B: Biological Sciences*, *266*(1433), 2131–2136.

Oudeyer, P.-Y. (2005). How phonological structures can be culturally selected for learnability. *Adaptive Behavior*, *13*(4), 269–280.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision research*, *46*(28), 4646–4674.

Perlman, M., Clark, N., & Johansson Falck, M. (2015a). Iconic prosody in story reading. *Cognitive science*, *39*(6), 1348–1368.

Perlman, M., Dale, R., & Lupyan, G. (2015b). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, *2*(8), 150152.

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in psychology*, *1*, 227.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *PNAS*, *108*(9), 3526–3529.

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.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, *26*(1), 43–49.

Sandler, W. (1996). Phonological features and feature classes: The case of movements in sign language. *Lingua*, *98*(1), 197–220.

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

Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt, Brace and company.

Schlehofer, D. (2016). *Sign Language Archaeology: Understanding the Historical Roots of American Sign Language*. Oxford University Press.

Schlenker, P., Chemla, E., Arnold, K., & Zuberbühler, K. (2016). Pyow-hack revisited: Two analyses of putty-nosed monkey alarm calls. *Lingua*, *171*, 1–23.

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). Major trends in vowel system inventories. *Journal of phonetics*, *25*(3), 233–253.

Scott-Phillips, T. C., & Blythe, R. A. (2013). Why is combinatorial communication rare in the natural world, and why is language an exception to this trend? *Journal of The Royal Society Interface*, *10*(88), 20130520.

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in cognitive sciences*, *14*(9), 411–417.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences*, *104*(18), 7361–7366.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423, 623–656.

Silvey, C., Kirby, S., & Smith, K. (2015). Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive science*, *39*(1), 212–226.

Stokoe, W. C. (1960). Sign language stucture: An outline of the visual communication systems of the american deaf. *Studies in Linguistics, Occasional papers, 8.*.

Studdert-Kennedy, M., & Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. In S. Kirby, & M. H. Christiansen (Eds.) *Language Evolution*, (pp. 235–254). Oxford University Press.

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature communications*, *7*.

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

Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*(16), 4530–4535.

Tria, F., Galantucci, B., & Loreto, V. (2012). Naming a structured world: a cultural route to duality of patterning. *PLoS ONE*, *7*(6), e37744.

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and cognition*, *4*(4), 357–380.

Verhoef, T. (2013). *Efficient coding in speech sounds: Cultural evolution and the emergence of structure in artificial languages*. Ph.D. thesis, The University of Amsterdam.

Verhoef, T., de Boer, B., & Kirby, S. (2012). Holistic or synthetic protolanguage: Evidence from iterated learning of whistled signals. In T. Scott-Phillips, M. Tamariz, E. Cartmill, & J. Hurford (Eds.) *The Evolution of Language: Proceedings of the 9th International Conference (EVOLANG9)*, (pp. 368–375). World Scientific.

Verhoef, T., Kirby, S., & Boer, B. (2015a). Iconicity and the emergence of combinatorial structure in language. *Cognitive science*.

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., & Padden, C. (2011). Cultural emergence of combinatorial structure in an artificial whistled language. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.) *Proceedings of the 33rd annual conference of the cognitive science society*, (pp. 483–488). Austin, Tx: Cognitive Science Society.

Verhoef, T., Roberts, S. G., & Dingemanse, M. (2015b). Emergence of systematic iconicity: Transmission, interaction and analogy. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. Jennings, & P. Maglio (Eds.) *The 37th annual meeting of the Cognitive Science Society (CogSci 2015)*, (pp. 2481–2487). Austin, TX: Cognitive Science Society.

Verhoef, T., Walker, E., & Marghetis, T. (2016). Cognitive biases and social coordination in the emergence of temporal language. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.) *The 38th annual meeting of the Cognitive Science Society (CogSci 2016)*, (pp. 2615–2620). Austin, TX: Cognitive Science Society.

Von Frisch, K. (1967). *The dance language and orientation of bees.*. Harvard University Press.

Waddington, K. D. (1982). Honey bee foraging profitability and round dance correlates. *Journal of comparative physiology*, *148*(3), 297–301.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, *128*(2), 179–186.

Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, *7*(03), 415–449.

Zipf, G. K. (1949). Human behavior and the principle of least effort: An introduction to human ecology.

Zuidema, W., & De Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, *37*(2), 125–144.