



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

SYSTEMATICITY, MOTIVATEDNESS, AND THE STRUCTURE OF THE LEXICON

Alan Nielsen

Thesis submitted for the degree of Doctor of Philosophy

School of Philosophy, Psychology, and Language Sciences

University of Edinburgh

2016



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- *This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.*
- *A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.*
- *This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.*
- *The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.*
- *When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.*

Abstract

For the majority of the 20th century, one of the central dogmas of linguistics was that, at the level of the lexicon, the relationship between words and meanings is arbitrary: there is nothing about the word ‘dog’ for example that makes it a particularly good label for a dog. However, in recent years it has become increasingly recognized that non-arbitrary associations between words and meanings make up a small, but potentially important portion of the lexicon. This thesis focuses on exploring the effect that non-arbitrary associations between words and meanings have on language learning and the structure of the lexicon. Based on a critical analysis of the existing literature, and the results of a number of experiments presented here, I suggest that the overall prevalence and developmental timing of two forms of non-arbitrariness in the lexicon— systematicity and motivatedness – is shaped by the pressure for languages to be learnable while remaining expressive. The effect of pressures for learnability and expressivity have been recognized to have important implications for the structure of language generally, but have so far not been applied to explain structure at the level of the lexicon.

The central claim presented in this dissertation is that features of the perceptual and cognitive organization of humans results in specific types of associations between words and meanings being easier for naïve learners to acquire than others, and that the pressure for languages to be learnable results in lexica that leverage these human biases. Taking advantage of these biases, however, induces constraints on the structure of the lexicon that, left unchecked, might limit its expressivity or penalize

subsequent learning. Thus, lexica are structured such that early-acquired words are able to leverage these biases while avoiding the limitations imposed by those biases when they are extended past a certain point.

Lay Summary

This dissertation focuses on the relationship between words and their meanings.

Typically, words are assumed to be related to their meanings based only on arbitrary convention: there is nothing about the word ‘tree’ that makes it a good word for a tall wooden plant, and thus it is not surprising that other languages use completely different words for trees. Some words however are not arbitrary: ‘oink’ is imitative of the sound that it describes, and can thus be described as motivated, rather than arbitrary. In addition to the possibility that words can be motivatedly connected to meanings, similar words like ‘glimmer’, ‘glitter’, and ‘glisten’ can refer to similar things: this type of non-arbitrary association between words and meanings is referred to as systematicity.

In this dissertation, I explore the effect that these two types of non-arbitrariness – systematicity and motivatedness – have on language learning. I suggest that languages, in order to be more learnable, take advantage of both of these types of non-arbitrary associations.

Acknowledgements

I want to thank, first, my supervisors, collaborators, and labmates who provided guidance and expertise over the period of my PhD. Thanks to Professor Simon Kirby and Dr. Jools Simner: both of you were instrumental in shaping the ideas in this dissertation and the path of my PhD Research. Dr. Kenny Smith, in addition to helping in those areas, thank you for all of the help with data analysis and working out the minutiae of experiments with me. All three of you have been enormously supportive through a stressful period and this dissertation wouldn't have been possible without you. I'd like to thank my labmates in the LEC, who provided support both social and academic that was vital to my growth as an academic. I'd especially like to thank Christine Cuskley for her advice and many great conversations, Matt Spike for continually scaring me about statistics, Cat Silvey for being the best cohort one could hope for, and Justin Sulik for providing incredible viticultural and gastronomic support, and being a good friend. In the environment of the University of Edinburgh more broadly, many other colleagues and professors have had a major impact on my academic development, and the fact that there are too many to list stands as testament to the research environment that I experienced during my PhD.

In addition to academic support, my journey to Edinburgh wouldn't have been possible without the enormous financial and emotional support provided by my

family and friends. My parents, Vagn and Linda Nielsen, despite not always agreeing with me about my ideas, fostered an environment that allowed me to grow intellectually and explore possibilities, and for that I owe them everything: I will one day repay them for their financial support, but will never be able to repay the debt of the opportunities that they have given me. To my brothers, Andrew and Adrian Nielsen, and their families – your support means everything. To my friends, both in Edinburgh and back home in Canada, thank you for keeping me sane and getting my head away from academic things when I needed to be reminded that there is a whole world out there worth exploring. Joseph Liptrap and Jamie Steer, especially, deserve thanks and appreciation for sticking with me as friends throughout my time in Edinburgh. Heriot's Rugby Club has been a second home to me in my time in Edinburgh, and reminded me that I have a body as well as a brain – COYN. To my friends back home, especially the Jameses, thanks for many late nights on skype that kept me rooted and feeling like I was never too far from home. To my partner, Christine, thank you for everything that you've done for me: you've kept me sane, fed, and smiling, and been a wonderful sounding board and advisor through the process of writing up. Additionally, your technical help in making figures for this dissertation helped not only to make it clearer for the reader but also to distill my own ideas.

Finally, thank you to the National Sciences and Engineering Research Council of Canada for a generous scholarship which helped me enormously in focusing on my

work, and also to previous researchers in my field. In academic writing we rarely get the time to thank those people on whose work our own is based- the fact that several names appear in this paper over 100 times makes clear how formative the ideas and methodologies of other researchers have been.

Declaration

I declare that the thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

The work presented in Chapter 3 includes a computational model created cooperatively between myself, Dieuwke Hupkes, and Dr. Kenny Smith^{**}. Dieuwke Hupkes was responsible for the creation of the model, under my supervision. The model results were analysed by myself. The remaining portions of Chapter 3 represent my own contribution.

The work presented in Chapter 4 has been submitted for publication in the *Journal of Experimental Psychology: Language, Memory, and Cognition* as Sound-symbolic words are easier to learn, but inhibit learning of conventional words by Alan Nielsen^{*}, Simon Kirby^{**}, Julia Simner^{**}, & Kenny Smith^{**}. The study was conceived primarily by me, Alan Nielsen, and I carried out the design and coding of the experiment, collection of data, and the majority of the statistical analysis and writing.

Alan Nielsen

May 1, 2016

***= Author of the Dissertation**

****= Supervisor**

Contents

1	Introduction and Thesis Overview	1
1.1	Motivatedness	6
1.1.1	Modality	8
1.1.2	Mechanisms	11
1.1.3	Limitations	12
1.2	Systematicity	15
1.2.1	Systematicity and Dimensionality	16
1.3	Systematicity and Motivatedness	23
1.3.1	Motivated Non-systematic	26
1.3.2	Motivated Systematic	29
1.3.3	Non-motivated Systematic	32
1.3.4	Non-motivated Non-systematic (Arbitrary)	36
1.4	Motivatedness and Language Learning	38
1.4.1	Empirical Evidence	38
1.4.2	Child-directed speech	40
1.4.3	Corpus analysis and cross-linguistic consistency	40
1.5	Systematicity and Language Learning	41
1.5.1	Experimental evidence	41
1.5.2	Computational models	43
1.5.3	Corpus analysis	43
1.6	Motivatedness, systematicity, and language learning	44
1.6.1	Bootstrapping	52
1.6.2	Contrastiveness and learnability	56
1.7	Thesis Outline	57
1.7.1	Overview and contribution statements	57

1.7.2	Chapter 2- Systematicity and language learning	58
1.7.3	Chapter 3- Phonology	59
1.7.4	Chapter 4- Motivatedness and Categorisation	59
1.7.5	Chapter 5- Growing Lexicon experiment	60
1.7.6	Conclusions	60
2	Systematicity and Learning I	62
2.1	Background and Rationale	65
2.2	Chapter Outline	68
2.3	Experiment 1	69
2.3.1	Methods	70
2.3.2	Results	79
2.3.3	Discussion	84
2.4	Experiment 2	90
2.4.1	Methods	91
2.4.2	Results	95
2.4.3	Discussion	100
2.5	Experiment 3	101
2.5.1	Methods	104
2.5.2	Results	111
2.5.3	Discussion	115
2.6	General Discussion and Conclusions	122
3	Systematicity and Learning II: Phonological Similarity	125
3.1	Background and Rationale	128
3.2	Chapter Outline	131
3.3	Simulation 1	134

3.3.1	Methods	134
3.3.2	Replication results	145
3.3.3	Replication discussion	148
3.3.4	Methods II	151
3.3.5	Results	154
3.3.6	Discussion	157
3.4	Experiment 4	159
3.4.1	Methods	161
3.4.2	Results	169
3.4.3	Discussion	173
3.5	General Discussion and Conclusions	174
4	Motivatedness and Categorisation						182
	Foreword	183
4.1	Background and Rationale	188
4.2	Investigating Motivatedness	193
4.3	Experiment 5	194
4.3.1	Methods	195
4.3.2	Results	203
4.3.3	Discussion	208
4.4.	Experiment 6	214
4.4.1	Methods	215
4.4.2	Results	222
4.4.3	Discussion	227
4.5	General Discussion and Conclusions	228
5	Growing Lexicon Experiment						236

5.1	Background and Rationale	237
5.1.1	Signal space saturation	238
5.1.2	Contrastiveness	240
5.1.3	Previous findings	242
5.1.4	Signal space saturation, contrastiveness, and learning	247
5.1.5	Sound-symbolic bootstrapping	249
5.2	Investigating signal space saturation and bootstrapping	252
5.3	Experiment 7	252
5.3.1	Methods	254
5.3.2	Results	279
5.3.3	Discussion	284
5.4	General Discussion	271
5.4.1	Extensions	287
5.5	Conclusions	291
6	Conclusions	294
6.1	Motivatedness, systematicity, and language learning	298
6.2	Summary	301
6.2.1	Experimental evidence	301
6.2.2	Bootstrapping	313
6.2.3	Contrastiveness and Confusability	314
6.3	Future Directions	316
6.3.1	Bootstrapping	316
6.3.2	Contrastiveness	319
6.4	Overall Conclusions	323
	References	326

List of Figures

1.01	The original ‘maluma’ and ‘takete’ shapes from Kohler (1947)	10
1.02	Conventionalisation of an iconic logograph. From Garrod et al. (2007)	14
1.03	The effect of systematic marking on the size of the signal space I	18
1.04	The effect of systematic marking on the size of the signal space II	19
1.05	The effect of systematic marking on the size of the signal space III	20
1.06	The crossing of systematicity and motivatedness	24
1.07	A representation of a motivated association	26
1.08	A representation of an incidentally systematic association	28
1.09	A representation of a motivated systematic association	30
1.10	An example of non-incidental motivated systematicity	31
1.11	An example of incidental motivated systematicity	32
1.12	A representation of a non-motivated systematic association	33
1.13	A representation of arbitrary associations	36
1.14	A simple model of the predicted structure of the lexicon	45
1.15	A more complex model of the predicted structure of the lexicon	47
1.16	A visualization of how systematicity constrains the signal space	51
2.01	A representation of the learnability pressures explored in Chapter 2	62
2.02	A representation of non-motivated systematic word-meaning mappings	63
2.03	d’ performance of participants in Experiment 1	80
2.04	Results of Experiment 1 comparing performance across trial types	82
2.05	Correlation of performance with contrastiveness in Experiment 1	84
2.06	Comparison of categorisation between Exp 1 and Monaghan et al.	87
2.07	Comparison of individuation between Exp 1 and Monaghan et al.	87
2.08	Comparison of true individuation between Exp 1 and Monaghan et al.	89

2.09	d' performance of participants in Experiment 2	95
2.10	Results of Experiment 2 comparing performance across trial types	97
2.11	Comparison of results between Experiment 1 and 2	99
2.12	Results of the 'Half-half' experiment from Monaghan et al. (201x)	103
2.13	d' performance of participants in Experiment 3	112
2.14	Results of Experiment 3 comparing performance across trial types	113
2.15	Correlation between performance and contrastiveness for Exp 3	115
3.01	A representation of the learnability pressures explored in Chapter 3	126
3.02	Average phonological feature edit distance for Replication	140
3.03	Average phoneme edit distance for Simulation 1	143
3.04	Comparison between Monaghan et al.'s model and my replication	147
3.05	Categorisation error comparison between Monaghan and replication	150
3.06	Phonological feature edit distance for phonemes in Simulation 1	152
3.07	Phonological feature edit distance for words used in Simulation 1	153
3.08	Results of individuation performance for Simulation 1	155
3.09	Results of categorisation performance for Simulation 1	156
3.10	Results of categorisation error for Simulation 1	157
3.11	A sample of the images used for Experiment 4	164
3.12	The average contrastiveness of the languages used in Experiment 4	167
3.13	d' performance from Experiment 4	170
3.14	Results of Experiment 4 comparing performance across trial types	172
4.01	A sample of stimuli from experiments exploring the Bouba-Kiki effect	190
4.02	An example of the figures used in Experiment 5	197
4.03	A diagrammatic representation of a sample trial from Experiment 5	201
4.04	Results of performance for Experiment 5	205

4.05	Performance on early trials for Experiment 5	.	.	206
4.06	Response time data for Experiment 5	.	.	208
4.07	Examples of images used in Experiment 6	.	.	216
4.08	Trial types from Experiment 6	.	.	220
4.09	Performance on trials where both size and shape are relevant for Exp 6			223
4.10	Performance on trials where only shape is relevant for Experiment 6			224
4.11	Performance on trials where only size is relevant for Experiment 6			226
5.01	The effect of systematicity on the signal space	.	.	239
5.02	Results of Individuation performance from Experiments 1 and 2			244
5.03	A sample training trial from Experiment 7	.	.	263
5.04	A sample testing trial from Experiment 7	.	.	264
5.05	Performance on Experiment 7 for each condition.	.	.	280
5.06	Summed confusability across 7 testing blocks for Experiment 7			281
5.07	Summed confusability for each trial type in Experiment 7			282
5.08	Performance on Experiment 7 including confusability as a predictor			283
6.01	A model of the pressures for learnability and expressivity.	.	.	298
6.02	Results from Experiments 1 and 2	.	.	303
6.03	Results from Experiment 4	.	.	306
6.04	Results from Experiment 5	.	.	309
6.05	Results from Experiment 6	.	.	309
6.06	Results from Experiment 7	.	.	311
6.07	Summed confusability metric from Experiment 7	.	.	312
6.08	“Yufu” stimuli from Lupyan et al. (2007)	.	.	318

List of Tables

1.01	The effect of systematicity on the size of the signal space	.	21
1.02	The overall size of a potential signal space	.	21
2.01	Word construction for Experiment 1	.	72
2.02	Comparison of words between Experiment 1 and Monaghan et al.		73
2.03	A comparison of words between Experiment 1 and Experiment 2		93
2.04	Signal space used in Monaghan et al.'s Half-half experiment	.	102
2.05	The creation of syllables for Experiment 3	.	106
2.06	The creation of words for Experiment 3	.	107
2.07	An example of a set of possible signal spaces for Experiment 3		108
2.08	Contrastiveness of systematic labels from Monaghan et al.	.	117
2.09	A sample language from Experiment 3	.	120
3.01	Phonological feature codings from Simulation 1	.	136
3.02	Phoneme sets for Simulation 1	.	139
3.03	Labels used in Simulation 1	.	142
3.04	A sample of possible languages for Experiment 4	.	163
4.01	Consonant phonemes used in Experiment 5	.	199
4.02	Sample words used in Experiment 5	.	199
4.03	Sample words used in Experiment 6	.	217
5.01	The four conditions used in Experiment 7	.	256
5.02	Consonant phonemes used in Experiment 7	.	257
5.03	A sample set of words for Experiment 7	.	258

5.04	The orthographic representation for words in Experiment 7	260
5.05	The number of words trained and tested in each block of Experiment 7	266
5.06	An example of calculating average contrastiveness from Experiment 3	267
5.07	The calculation of the confusability metric for Experiment 7	269
5.08	Alternative confusability metrics that were not used for Experiment 7	269
5.09	Trial types for Experiment 7	278

Chapter 1

Introduction and Thesis Overview

How do words get their meanings? This question, which Harnad called the symbol grounding problem (1990) dates back to at least Plato's Cratylus dialogue, which is recognized as the first recorded exploration of the issue. Generally, linguistic tradition has been built around the idea that there are no connections between words and meanings other than those established by linguistic convention (de Saussure, 1983; Hockett, 1960). In other words, the lexica of natural languages are arbitrary when it comes to how the form of a word (i.e., its sound) is related to its meaning. For example, there is nothing about the word 'tree' that makes it particularly good for describing a large plant made of wood, and thus different languages have different words for trees. In recent years, however, the proposal that lexica are entirely arbitrary and conventional has come under close scrutiny. Word-meaning mappings can in fact be non-arbitrary in two ways, and researchers have increasingly recognized that these non-arbitrary associations can be found in the lexica of natural languages. First, relationships between individual words and their meanings can be non-arbitrary because they are *motivated* by the perceptual and cognitive organization of language users (Dingemanse et al., 2014): for example, the word 'oink' is imitative of its meaning (the sound that a pig makes). Second,

associations between words and meanings can be non-arbitrary by virtue of being *systematic* (Monaghan et al., 2014). Systematic relationships between words and meanings refer to a configuration where similar words are mapped onto similar meanings: ‘glimmer’, ‘glitter’, and ‘glisten’ for example are similar to one another both because they all begin with the segment ‘gl-’, and because they all have meanings related to light.

The goal of this thesis will be to explore these non-arbitrary associations between words and meanings, evaluating and providing evidence for the proposal that non-arbitrary relationships between words and meanings have important influences on learning and the structure of the lexicon. The basic assertion proposed here is that the pressure for languages to be learnable (Kirby et al., 2015) is met by taking advantage of non-arbitrary (motivated or systematic) associations between words and meanings. Because both motivated (Section 1.1) and systematic (Section 1.2) non-arbitrary associations between words and meanings are easier to learn than arbitrary associations (Sections 1.4-1.5), the process of cultural evolution should result in lexica that take advantage of these associations.

The most basic prediction that can be made by invoking the pressure for languages to be learnable is that the lexicon should be primarily non-arbitrary, because both systematic and motivated associations between words and meanings are easier to learn than arbitrary ones. This prediction, however, is not borne out in

the lexica of natural languages, which are largely arbitrary. The central aim of this dissertation is to account for the overall arbitrariness of the lexicon while allowing non-arbitrary associations to contribute meaningfully to the process of language learning. Specifically, I will suggest that both motivatedness and systematicity have inherent limitations that preclude them from making up the majority of the lexicon, but that the developmental time course over which human learners acquire words can allow for those non-arbitrary connections to have an important influence on learning when their limitations are less robust.

The experimental results presented in this dissertation allow for a more complete explanation of this possibility than has been previously undertaken (Monaghan et al., 2011; Monaghan et al., 2014). Primarily, this dissertation explores how the pressure for learnability (languages must be learnable) and the pressure for expressivity (languages must be able to express a sufficient number and range of concepts) interact with human cognitive and perceptual biases over the developmental time course of language acquisition to shape the structure of the lexicon.

It is important to note from the start that this dissertation focuses on non-arbitrariness at the level of the lexicon, focusing on monomorphemes and pseudoword stimuli. At the level of morphosyntax, it is uncontroversial that a great deal of human language is systematic (e.g. plurality in English is morphologically

marked, typically by adding /-s/), but at the level of the lexicon, language is assumed to be arbitrary. Nonetheless in several places in this dissertation, I will return to a discussion of non-arbitrariness in morphosyntax, especially insofar as the pressure for learnability that has shaped morphosyntax might inform our search for similar processes in the formation of the lexicon.

This chapter will focus on a review of the literature surrounding non-arbitrariness in the lexicon, providing a necessary standardization of the terminology used in studies ranging both across disciplines (primarily psychology and linguistics, but also philosophy, anthropology, and behavioural ecology) and across time (from the early twentieth century to more contemporary research). In section 1.1, I explore motivated associations between words and meanings, and suggest that the term motivatedness encompasses a wide swath of concepts in language learning. In this discussion of motivatedness I survey a broad range of such associations before discussing the likely mechanisms that underpin those associations and thus the suggestion that they can be co-opted for learning in language-naïve learners. Finally, I discuss the limitations of motivatedness, suggesting that motivated associations between words and meanings can express a finite number of concepts that are insufficient for an expressive language.

In section 1.2 I explore systematic mappings between words and meanings and examples of those types of associations in the lexicon. I also provide a thorough

discussion of the limitations of systematic mappings between words and meanings, suggesting that systematicity can constrain expressivity and also potentially learning - an issue that will become central to the rest of the dissertation.

In section 1.3 I discuss the intersection of systematicity and motivatedness, and suggest that the two types of non-arbitrariness are orthogonal to one another: words can be non-arbitrarily related to their meanings both as a function of being motivated and as a function of being systematic at the same time. Thus, connections between words and meanings can be motivated but non-systematic, motivated and systematic, non-motivated but systematic, or non-motivated and non-systematic (arbitrary). I provide examples from natural languages that combine features of the two types of non-arbitrariness, and discuss their relative contributions to the lexicon.

In sections 1.4 and 1.5 I return to discussions of motivatedness and systematicity respectively, focusing on the influence that the presence of these types of non-arbitrary associations has on learnability. To this end, I consider evidence from experimental studies, corpus analyses, and computational simulations of learning.

In section 1.6 I return to the suggestion that the lexicon should be non-arbitrary, incorporating the evidence discussed in the preceding sections to inspire a more complex model for the structure of the lexicon that considers pressures for the

language to be expressive, and also the possibility of learnability pressures favoring arbitrariness. This more complex model incorporates existing findings in the literature, but also makes clear what conceptual and evidentiary pieces are missing. The search for support for this model becomes the central drive of this dissertation.

Finally, in section 1.7 I outline the structure of the remaining chapters of this thesis with reference to the central questions suggested by the model proposed in section 1.6.

1.1 Motivatedness

In this section I will discuss motivatedness, which refers to a configuration of language where some feature of a word is mapped onto a related feature of its meaning. I will begin by differentiating between unimodal and crossmodal associations, providing examples of each, before considering the mechanistic explanations underlying those associations and their limitations for expressing dimensions relevant to language.

The concept of relatedness, that is, the dimension along which a feature of a word can be mapped to a feature of its meaning, is quite broad. First, motivated mappings between words and meanings can vary in terms of their modality: for example, the onomatopoeic word “oink” is similar to the sound that it describes, while the association between the word “teeny” and a small object is mediated crossmodally by virtue of the perceptual relatedness of high pitched sounds and

small objects (Sapir, 1929). Motivatedness as a broad term captures all possible non-arbitrary relationships between features of words and features of meaning while acknowledging that these relationships can be mediated along a number of dimensions.

Despite the fact that language at the level of the lexicon is typically thought of as arbitrary, recent research has suggested that motivatedness is an important property of all languages, whether they are signed or spoken (Perry, Perlman, & Lupyan, 2015). However, as I have acknowledged, motivatedness takes a number of forms: words can be related to meanings directly and unimodally (e.g. ‘oink’), or based on mappings between seemingly unrelated crossmodal dimensions (e.g. ‘enormous’ for large objects (Cuskley, 2013)). In both of these cases, the transparency and strength of the motivatedness of the relationships can also vary (‘moo’ is more similar to the sound of a cow than ‘cock-a-doodle-doo’ is to the sound of a rooster), both as a function of the modalities along which those relationships are structured and as a function of the language in which they are embedded. Finally, the mechanisms underlying these relationships might vary as a function of these features as well, with some associations being strongly biased by the perceptual or cognitive system of the language learner and others less strongly so, or contingent on the presence of contextual information. In the proceeding sections, I will discuss each of these topics and how they relate to motivatedness.

1.1.1 Modality

Motivated associations between words and meanings can be mediated by associations that vary in their modality. Broadly, associations can be either unimodal or crossmodal. In unimodal associations, a feature of meaning is mapped to a feature of the word along a dimension that is shared between the two (e.g. ‘crash’ is imitative of the sound of something crashing). Crossmodal associations are a type of motivated connection that relies on biases in cognition or perception that link otherwise unrelated features of words and meanings to one another: for example, humans have a bias to associate high pitch with small size, and this can be realized linguistically by the use of words like ‘teeny’ for small objects.

Unimodal associations

Typically, research into language focuses on unimodal associations based on acoustical properties (sound to sound) in spoken languages (Cuskley, 2013; Imai & Kita, 2014), and iconic representations of meaning features like size, shape, or movement in signed languages (Taub, 2001; Perniss & Vigliocco, 2014). One of the most widely discussed and commonly occurring unimodal associations is onomatopoeia, which can be found in many languages. In onomatopoeia, the non-linguistic sound is mapped to a word that is in some way imitative of that sound, for example, the word “woof to describe the sound that a dog makes. These associations

are often referred to as *iconic* (the structure of a word resembles what it stands for: Ahlner & Zlatev, 2010; Perniss & Vigliocco, 2014; Dingemanse et al., 2015) in the psycholinguistic literature.

Crossmodal associations

Crossmodal associations between words and meanings come in a variety of forms that attest to the possibility that crossmodal perception might be a major feature of human cognition. Because languages are typically transmitted acoustically, the types of crossmodal associations seen in language are often mediated by motivated associations between the sound of a word and some feature of its meaning, and thus I will focus on those types of associations.

Shape-sound symbolism

One of the most widely attested crossmodal associations that can be observed in humans is shape-sound symbolism. The Bouba-Kiki effect, which is the most well studied form of sound symbolism in the psycholinguistic literature was first discovered by Kohler (1929), who demonstrated a perceptual bias wherein experimental participants associated pseudowords with certain phonological features to objects that varied in their overall shape (Figure 1.01).

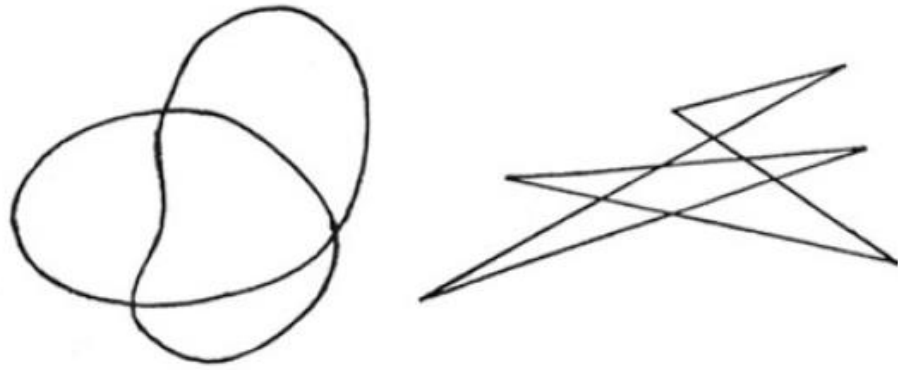


Figure 1.01- *The original 'maluma' (left) and 'takete' (right) shapes from Kohler (1929).*

The bias underlying the Bouba-Kiki effect is robust, having been demonstrated in both children (e.g. Ozturk, Krehm, & Vouloumanos, 2013) and adults, as well as with speakers of multiple languages (Davis, 1961; Bremner et al., 2013).

Size-sound symbolism

The Bouba-Kiki effect is perhaps the most well attested crossmodal linguistic bias experimentally, but size-sound symbolism is more widely attested crosslinguistically. Beginning with Sapir (1929), a number of authors have found pervasive relationships between vowel height and size, such that high front vowels are associated with small objects and low back vowels with larger objects. This effect has been attested in not only speakers of English (Johnson, 1967), but also Chinese,

Thai (Huang, 1969), Korean (Kim, 1977), and several other languages (Gebels, 1969; Malmberg, 1964; cf Newman, 1933, Newmeyer, 1993).

1.1.2 Mechanisms

Unimodal associations between words and meanings do not seem to require a very complex mechanistic explanation. For production, these associations require language users be able to produce signals that are recognizably similar to their meanings, taking into account phonotactics constraints. The recognition of the motivated relationship between these types of associations also seem straightforward, but less transparent word-meaning mappings might actually require fairly advanced cognition (section 1.13).

Some observed crossmodal biases seem to be best explained by simple associative learning mechanisms. For example, the association between distance and amplitude is a reliable feature of the environment that can be learned from the environment and then leveraged later for communication (distant sounds are, all things being equal, quieter). Other associations, like for example associations between high pitched sounds and perceptual brightness, seem less obvious and might require some other mechanistic explanation.

Crossmodal associations reflect a general feature of human perceptual organization where cortical areas for different modalities are connected (e.g. Kovic,

Plunkett, & Westermann, 2010). Evidence for crossmodal associations being underpinned by structural organization of the brain is found primarily in the study of synesthesia, a cluster of conditions that results in exaggerated connections between unrelated sensory modalities (Ramachandran & Hubbard, 2001). Many, but not all crossmodal associations observed in synesthetes have also been demonstrated in the normal population, although synesthetes might show exaggerated versions of those associations (Bankieris & Simner, 2015).

The physiological and genetic (Asher et al., 2009) underpinnings of crossmodal associations in both synesthetes (Rouw & Scholte, 2007) and the normative population (Revill et al., 2014) have been shown to be related to one another. Associations across sensory modalities have been demonstrated to be mediated by increased density of neural connections between the cortical areas responsible for the processing of those inputs in both synesthetes and the normal population (Kanero et al., 2014).

1.1.3 Limitations

The acoustic channel on which most languages are transmitted inherently limits the possibility for unimodal associations between words and meanings: meanings having to do with imitable acoustic dimensions can be matched to labels that are motivated, but the majority of meanings necessary for human language are simply not amenable to this kind of unimodal motivated mapping. Human languages need

to express things about the world other than descriptions or imitations of sounds. Additionally, motivated associations between words and meanings are not equally transparent: some meanings, for example, map relatively poorly onto their communication channel (i.e., whether the language is verbal, signed, written, etc.). The transparency of unimodal motivated associations is also related to the process of conventionalisation: although unimodal words like onomatopoeia are imitative of sounds, they are constrained by the phonotactics of the language in which they are embedded.

In English, for example, the word for the sound of a crowing rooster ('cock-a-doodle-doo') is onomatopoeic. Other languages, however, have different expressions for this sound: German uses 'kikiriki', where French uses 'cocorico'. I might, as an English speaker, suggest that the English onomatopoeia is more straightforwardly iconic than the other two, but all three are quite different from the actual sound made by a rooster (Perniss, Thompson, & Vigliocco, 2010). Human language, especially when conventionalized to recognizable words rather than non-speech sounds is unable to perfectly mimic the sounds of acoustic events, because onomatopoeia are still formed by a string of speech sounds (e.g. vowels and consonants), and, even if phonemes could perfectly mimic environmental sounds, all languages have constrained sets of available phonemes (Assaneo et al., 2011).

Crossmodal associations between words and meanings have similar limitations based on the suitability of the communication channel for taking advantage of perceptual biases. In some ways, because these associations are not imitative, the fidelity of imitation is not an issue, but at the same time the fidelity of recognition might be more difficult. Absent prosodic cues, for example, it might be difficult to recognize that the vowels in ‘humongous’ make it an appropriate word to describe very large things.

The Chinese symbol for ‘gate’ serves as a good example of the process of conventionalisation of a motivated word-meaning association, in this case demonstrating the process of erosion. Over time the initially unimodally iconic representation of a gate in Chinese became increasingly arbitrary and divorced from its iconic origin (Figure 1.02). This further obscures the motivated nature of the association between the logograph and its meaning.



Figure 1.02- The conventionalisation of an iconic logograph for the word GATE in Chinese.

From Garrod et al. (2007). p. 962.

Crossmodal associations are likely subject to the same processes of erosion and conventionalisation as unimodal associations. Dingemanse (2012) has suggested, for

example, that the process of conventionalisation might result in motivated crossmodal associations that are language specific, although this possibility seems to require that these iconic associations are also systematic, or at least derived from the structure of a language more generally.

Collectively, motivated associations between words and meanings are potentially beneficial for language learning, but limited in their expressivity. This limit to expressivity has both an absolute component (some meanings do not have motivated associations to a feature of the communication channel) and a relative component based on transparency (the motivatedness of certain associations might not always be entirely straightforward).

1.2 Systematicity

Where motivatedness refers to associations that operate directly between words and their meanings, the second non-arbitrary dimension along which language can be structured operates by mapping characteristics of sets of words to characteristics of sets of meanings (systematicity). In this section I will outline the effect that systematic associations between words and meanings has on the dimensionality of languages and thus their potential ability to express a sufficient number of meanings.

In an alien language where all proper nouns end with ‘-iks’, for example, any time a learner came across ‘-iks’ they would know immediately that they were

dealing with a proper noun. This sort of mapping, where all meanings of a certain category are mapped to all words with a given feature (and no words with that feature are mapped to other categories of meaning) might be called absolute systematicity, although generally (perhaps always, in natural languages) systematicity operates at a statistical level, rather than an absolute one (Reilly et al., 2012). Crucially, systematicity is orthogonal to motivatedness: that is, it is possible for the systematic mappings between word set features and meaning set features to be either motivated or conventional and idiosyncratic to a language (see section 1.3).

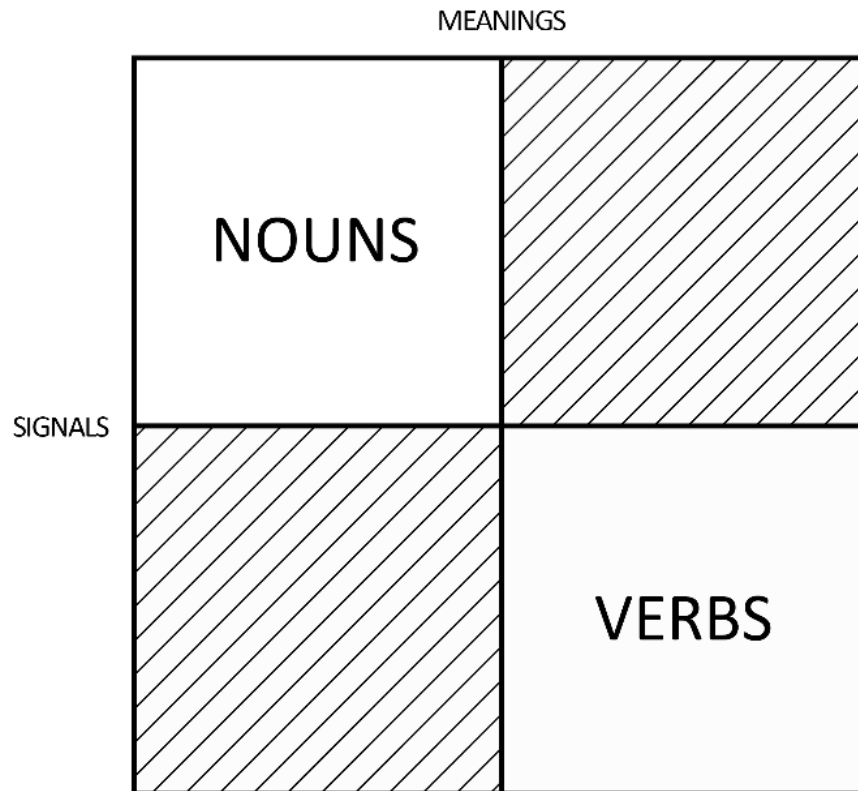
1.2.1 Systematicity and Dimensionality

One of the most important effects of creating systematic associations between properties of word sets and properties of meaning sets is that doing so inherently limits the dimensionality of the signal space for a language. This is immediately apparent in the example above of absolute systematicity: because the ending ‘iks’ can only be used for proper nouns in such a configuration, the overall expressivity of that language has been reduced (there are many possible words that cannot be used for proper nouns).

Quantifying the dimensionality of language, however, and especially quantifying the loss of potential expressivity based on the introduction of a systematic division to the signal space, is not easy. How the introduction of

systematic divisions influences the potential expressivity of a language varies as a function of the number of phonemes, the allowable length of words, and the number of systematic divisions required. Even if we assume some incredibly large value for the number of possible phonemes, or the length of allowable words, the introduction of systematic divisions always reduces the dimensionality of a language, and might as a corollary limit its expressivity.

The World Atlas of Language Structures (Dryer and Haspelmath, 2013) lists !Xóõ as the language with the largest inventory of consonant phonemes, at 122 (Maddieson, 2013) and German as the language with the largest inventory of basic vowels, at 14 (Maddieson, 2013b). A language using these consonant and vowel phoneme inventories (which would be the language with the largest basic phoneme inventory), and whose words were only cVc trigrams would have a possible signal space with over 200,000 words (an average language, by comparison, with 22.7 consonants and 6 vowels would have just over 3000 possible trigrams). If, in this language, we wanted to mark a distinction between nouns and verbs systematically we might suggest that all words for nouns should begin with half of the possible consonant phonemes, and all words for verbs should begin with the other half. This would result in just over 50,000 possible mappings each for nouns and verbs (1/4 of the total space each), which still seems like plenty of space, regardless of the fact that half of the previously available mapping space is lost by the introduction of the systematic marker (Figure 1.03).



***Figure 1.03-** The introduction of a systematic division marking nouns vs. verbs constrains the available mapping space. Areas of the space filled by hash marks represent mappings of words to meanings that are disallowed based on the systematic marking.*

The introduction of additional systematic divisions beyond a distinction between nouns and verbs further constrains the mapping space. If we introduce an additional marked distinction between count nouns and mass nouns, for example, we further reduce the number of possible labels for count nouns to 12,500 (Figure 1.04).

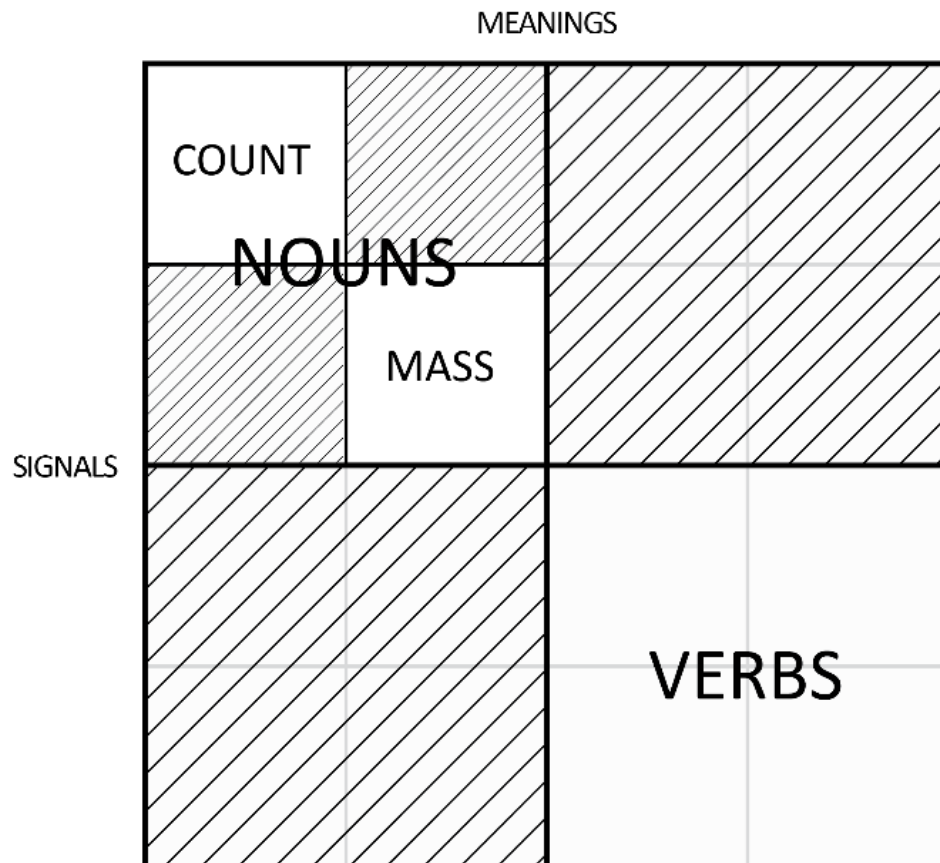


Figure 1.04- The introduction of a second systematic division splitting nouns into count vs. mass nouns further constrains the available signal space.

Further divisions between subtypes of nouns that are marked systematically will, of course, further constrain the usable mapping space (Figure 1.05). With the introduction of a third systematic division (animate vs inanimate count nouns) there are only 3000 possible labels for animate count nouns.

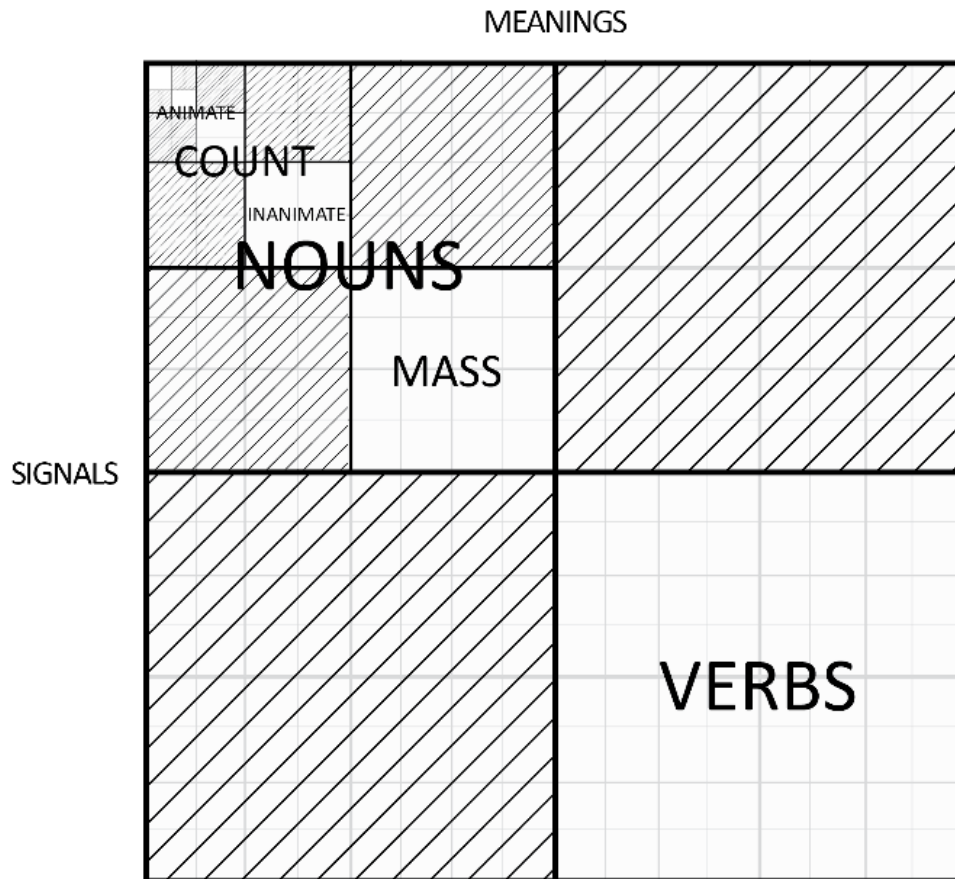


Figure 1.05- The introduction of a second systematic division splitting nouns into count vs. mass nouns further constrains the available signal space.

Of course, we could imagine further subdivisions of the available mapping space that would constrain the number of possible words in our artificial language. With each division, the number of possible words for a given meaning becomes exponentially smaller, and at some point might become so small that there would be an insufficient number of possible words to express all of the meanings within a category (Table 1.01).

<u>Level of Systematicity</u>	<u># of Possible Words</u>		<u>Percentage of unavailable mapping space</u>
	<u>Large Phoneme Inventory</u>	<u>Average Phoneme Inventory</u>	
Arbitrary	200000	3100	50
Noun/Verb	50000	773	75
Count/Mass	12500	193	87.5
Animate/Inanimate	3125	48	93.8
Natural/Artificial	781	12	96.9
Animals/Non-Animals	195	3	98.4
Domesticated/Wild	49	0.75	99.2
Pet/Farm Animal	12	0.19	99.6

Table 1.01- *The effect of systematicity on the size of the available mapping space. At each added level of systematicity, an increasingly large percentage of the mapping space becomes unavailable for use, and the total number of possible words within subcategories becomes smaller. At some level of systematicity, the number of possible words within a category becomes insufficient to express the meanings required for the language.*

Natural languages are typically more constrained in the size of their phoneme inventory than our artificial language presented above, but are less constrained in the way that words can be created (not all words are trigrams). The overall size of the possible signal space for any language is thus very large. Consider table 1.02 below, where the number of possible words of each length given our language with a large phoneme inventory vs. an average inventory.

	Consonants	Vowels	cV words	cVc words	cVcV words	cVcVc words	cVcVcV words
Large Inventory	122	14	1,708	208,376	2,917,264	355,906,208	4,982,686,912
Average Inventory	23	6	136	3,092	18,550	421,095	2,526,570

Table 1.02 *The overall size of the possible signal space for a language increases exponentially as a function of the total phoneme inventory and the allowed length of (monomorphemic) words.*

Here it might become obvious that the size of a language's signal space and its mapping space are not identical to one another. It would be possible, for example, to have a language with a very large signal space (many possible words), but a very constrained mapping space. A language with a small signal space and no systematic divisions, for example, might actually have more possible word-meaning mappings than a much larger language that incorporated many systematic divisions to the mapping space. For artificial examples like the one given above then, the absolute signal space of the language does not change as a function of introducing systematicity: discounting phonotactics and other rules for word construction, all words in the overall signal space are theoretically available for the language. Crucially, however, the introduction of systematicity ensures that certain mappings between words and meanings will not be available. Assuming an equal number of word-meaning pairs for a language that is either systematic or not, we would say that both languages make use of an equal proportion of the signal space. The crucial difference is that a systematic language maps from the signal space to the meaning space in such a way that large portions of the mapping space are unavailable – a constraint which is not shared by arbitrary languages.

It might be hard to imagine, given the nearly 5 billion possible trisyllables in our artificial example, that the introduction of systematicity might materially influence the expressivity of such a language, but consider that there are many more

differences than, for example, nouns vs. verbs that are relevant to language learners. The question of the degree to which systematicity imposes actual limits on the expressivity and learnability of language is central to much of this dissertation, and is dependent on the size of the signal space, the number of systematic dimensions along which the language is to be marked, and how those two features interact with the perceptual and cognitive machinery of language learners.

1.3 Systematicity and Motivatedness

As mentioned above, systematicity and motivatedness are not mutually exclusive and, in fact, run orthogonally to one another – that is, the relationship between a word and meaning can be systematic vs. non-systematic and motivated vs. non-motivated. There are examples of all of these possible configurations in natural languages, although non-motivated non-systematic (i.e. arbitrary) associations between words and meanings account for by far the largest portion of the lexicon (de Saussure, 1983). In the sections below I will provide examples of each of these types of associations between words and meanings. Additionally, taking into consideration the benefits of systematicity and motivatedness and their limitations, I attempt to explain the relative contribution of each type of associations to the lexicon.

Motivated Systematic	Motivated Non-Systematic
<p>A part of the language where there are both mappings between individual words and their meanings (motivatedness) and between groups of words and groups of meanings (systematicity).</p> <p>Examples: Pervasive sound symbolism Some phoneasthemes (e.g. gl-cluster)</p>	<p>A part of language where there are mappings between individual words and their meanings (motivatedness) but no relationship between groups of similar words and groups of similar meanings.</p> <p>Examples: Onomatopoeia Mimetics/Ideophones</p>
Non-Motivated Systematic	Non-Motivated Non-Systematic
<p>A part of language where there is no mapping between individual words and their meanings, but where features of a set of similar words are mapping onto a set of similar meanings (systematicity).</p> <p>Examples: Phonaesthemes</p>	<p>A part of language that is arbitrary. This makes up the majority of the lexicon.</p>

Figure 1.06- The crossing of systematicity and motivatedness creates four possible ways that words can be related to meanings, which should account for the entire lexicon.

Before proceeding into a more complete discussion of the relationship between motivatedness and systematicity, it bears mentioning at least two things that make the practical distinction between the two-types of non-arbitrariness fuzzier than the treatment that I give them below. First, recognizing systematicity can be problematic: this is true not only because systematicity is most often statistical

rather than absolute, but also because there is no hard and fast cut-off point for systematicity. If a language has, for example, two words for small objects that contain a specific vowel sound could this rightly be described as systematicity? What about three words, or four? Making a call here is difficult, and when looking for systematicity we should be mindful of what kinds of evidence we accept, lest we have a very high false positive rate for systematicity, and report systematicity (either absolute or statistical) that is illusory.

Relatedly, we should recognize that the relationship between motivatedness and systematicity can be a slippery one. Below, I introduce the idea of *incidental systematicity*, which refers to configurations where by taking advantage of motivated associations repeatedly the language arrives at a configuration that could also be described as systematic. In the above example, consider the possibility that words for small objects contained high-front vowels – an association we know to be motivated (Sapir, 1929). Just as it can be difficult to establish whether this configuration was systematic, it might be hard to find cases where motivated associations are taken advantage of in a language that does not result in incidental systematicity. This is further complicated by the fact that whether we recognize a group of words as being systematic depends on the level at which they are considered. For example, ‘moo’, ‘tweet’, and ‘oink’ are all motivated, but neither the phonological features of the words are similar nor are the acoustical features of the sounds that they represent, so we might call them non-systematic. At the same time,

however, all three words are similar in that they are imitative of the sounds of animals – should we consider the broad use of onomatopoeia in English an example of systematicity? If so, is it incidental or not?

These two considerations muddy the water of the discussion below, and should be kept in mind, but for the majority of the remaining dissertation I discuss more ideal less complicated examples of non-arbitrariness.

1.3.1 Motivated Non-Systematic

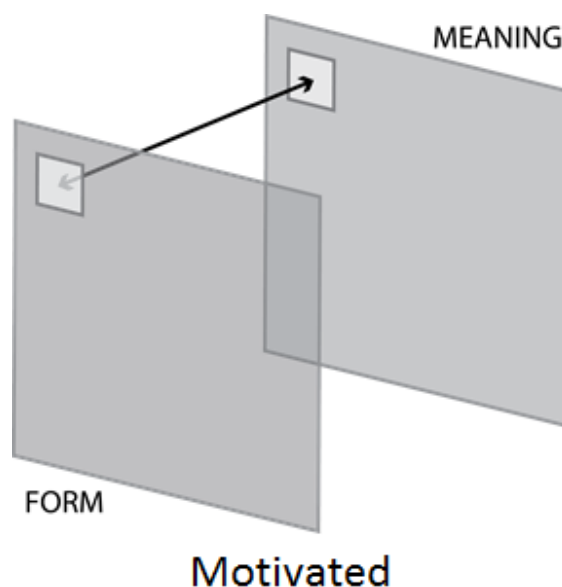


Figure 1.07- A diagrammatic representation of a motivated non-systematic association between a word's form and its meaning.

Motivated non-systematic associations are those where there is a mapping between some feature of a word's meaning and some feature of the word. In English,

onomatopoeia is an example of a motivated non-systematic association, while ideophonic forms in other languages also meet this criteria; ‘goron’, for example, is sound-symbolically associated with heaviness in Japanese (Asano et al., 2015), likely due to its vowel roundedness. Here I propose that motivated non-systematic associations between words and meanings likely make up the smallest portion of the lexicon for two reasons. First, despite the fact that there are many well attested crossmodal perceptual biases that might underpin these kinds of associations, there is still a limit to the number of concepts that can likely be expressed along a motivated dimension. We might consider, for example, the form of the words ‘lullaby’ and ‘fuck’ to be related to their meanings motivatedly: ‘lullaby’ is more sonorous than is ‘fuck’, and this difference maps onto their underlying affective dimensions (Yardy, 2010). Words like ‘lullaby’, and ‘fuck’ might thus be particularly well suited for their meanings, but what might a motivated signal for ‘honor’ or even ‘signal’ be? The second reason that I propose that isolated motivated non-systematic associations between words and meanings might account for a very small portion of the lexicon is that the introduction of multiple signals based on a single motivated dimension inherently produces mappings that are systematic in addition to being motivated.

Curse words like ‘fuck’, for example, use less sonorant consonants and prosodic cues that enhance their spectral harshness, and this has been suggested to be related to the motivational, hedonic, and affective purpose of those words

(Yardy, 2010). In a language with only a single curse word, the use this association would be only motivated and not systematic. However, in English and most other languages we have a large repertoire of curse words, and insofar as each of these words matches to the motivated bias we suggest, the structures of sets of meanings and sets of signals begin to come into an alignment and gain the property of being systematic (see Figure 1.08).

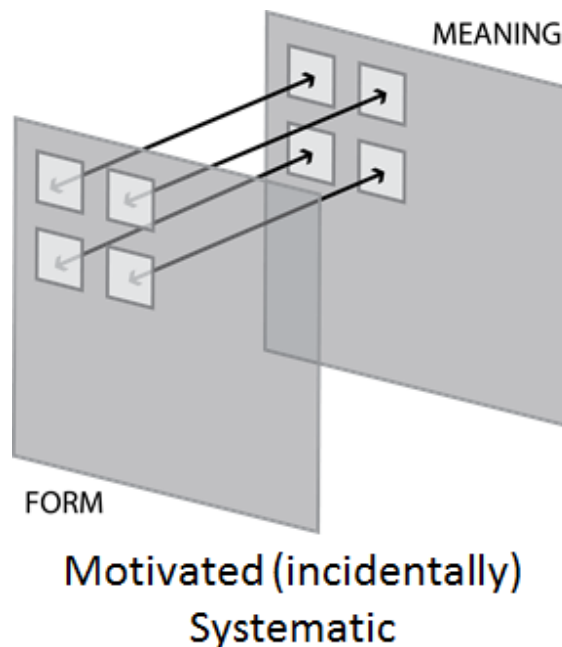


Figure 1.08- A diagrammatic representation the formation of an incidentally systematic mapping between words and meanings. In this case, individual word forms are mapped to meanings based on a motivated association between the two. Because multiple similar word forms are mapped to multiple similar meanings, the resulting configuration of the language can be recognized as being systematic.

1.3.2 Motivated Systematic

In this dissertation, I call the above shift from a motivated non-systematic association to a set of motivated systematic associations *incidental systematicity*, which could account for some observed types of word-meaning associations in natural languages. In English for example, the border between phonaesthemes (clusters of similar words that are mapped to similar meanings) and onomatopoeia can be a blurry one: onomatopoeic words like ‘crash’, ‘clang’, ‘smash’, ‘bang’, and ‘crunch’ express similar meanings and are also similar to one another (in that they are similarly structured and use overlapping segments). Thus, the motivatedness of these form-meaning mappings also creates an overall structure that is recognizably systematic. Some ideophonic expressions in other languages share this feature. For example, reduplication is a common feature in Japanese ideophones that describe events occurring repeatedly (Asano et al., 2015), and is suggested to be motivated, but because the same motivated mapping is used repeatedly, those clusters of ideophones are also more systematic than would otherwise be expected.

The non-incidental case of motivated systematic mappings between words and meanings might be one where there is a motivated mapping between a property of a set of words and a set of meanings, but where no such motivated mapping can be recognized in the comparison of any individual word to its meaning (Figure 1.09). Dingemanse (2011) outlines the existence of this kind of association in Siwu where a

group of words that varies in their vowel quality (‘pɔmbɔlɔɔ’, ‘pumbuluu’, and ‘pimbilii’) maps to meanings about the protrusion of the belly, with /ɔ/ being mapped to the largest protrusion and /i/ to the smallest.

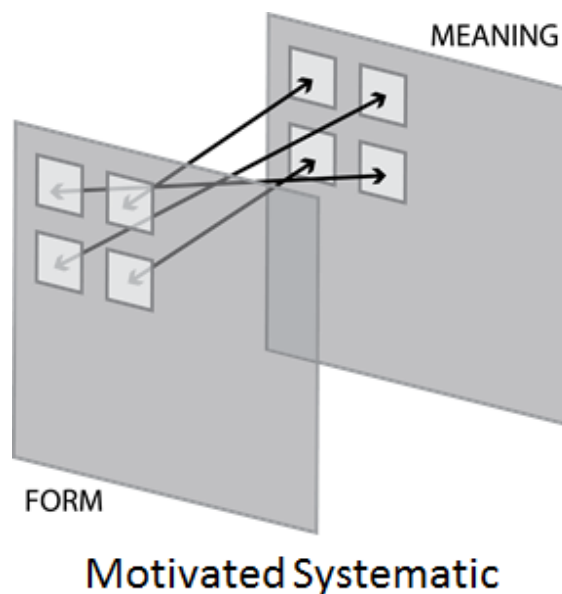


Figure 1.09- A diagrammatic representation the association between word forms and meanings that is both motivated and systematic. In this case, the motivated connection between the form of a set of words and the form of its related set of meanings is based not on individual motivated associations but a motivated mapping of the set of word forms to the set of meanings.

As an example, consider the names for a fictional family of animals that includes three species that differ primarily in how dangerous they are: we might have a domesticated cat; a wild cat, but one that is not typically thought of as being dangerous (e.g. a bobcat); and a large, wild, and very dangerous variety (e.g. a tiger). For these meanings, there is a dimension (danger) that increases across the group,

and this dimension could be mapped onto an analogous signal dimension that similarly increases (see Figure 1.10).

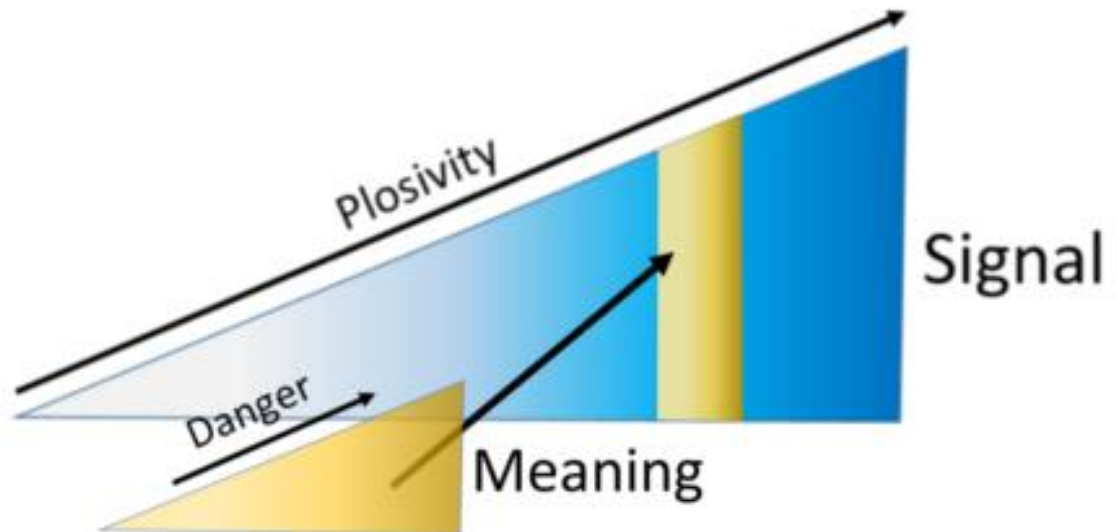


Figure 1.10- An example of non-incidental motivated systematicity. Here, a similarity on the meaning dimension (increasing danger) is mapped onto a similar structure on the meaning dimension (plosivity). In this case, the low danger meanings are mapped onto relatively less plosive signals, whereas the more dangerous meanings are mapped onto relatively more plosive signals.

In the example above, the mapping between the signal space and the meaning space is motivated in the normal sense (i.e. there is a motivated affective connection between plosivity and danger), but the individual mappings between words and meanings might not themselves be sufficiently transparent to be recognized as iconic. That is, the entire meaning dimension is mapped only onto a relatively small portion of the signal dimension. This mapping to the signal space might be so

narrow as to be insufficient for any individual word to be iconic, and thus the motivatedness of this systematic relationship arises as a function of the similarity solely between the set of meanings and the set of signals. Incidental motivated systematicity, on the other hand, would maintain individual iconic associations, with the systematicity arising because of the relationship between those associations (Figure 1.11).

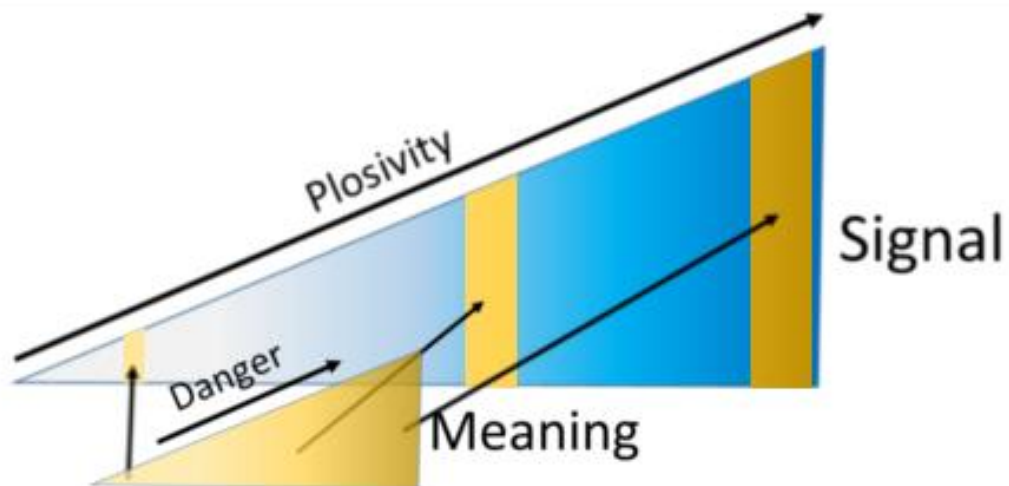


Figure 1.11- An example of incidental motivated systematicity. Here meaning dimensions are mapped onto signal dimensions in a motivated way individually. However, this mapping incidentally also creates a systematic structure to the relationship between the set of meanings and the set of signals.

The difference between incidental and non-incidental motivated systematicity is thus quite subtle, and the two are likely to bleed into one another: i.e. non-incidental systematicity might make the motivatedness of associations between individual words and meanings more obvious.

1.3.3 Non- Motivated Systematic

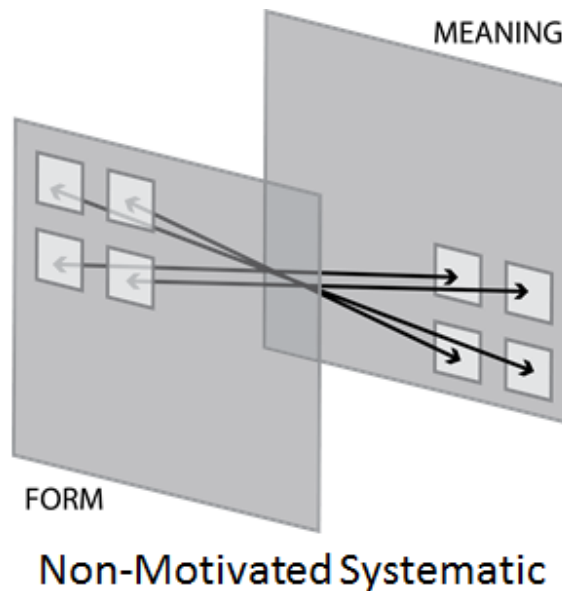


Figure 1.12- A diagrammatic representation of a non-motivated but systematic association between word forms and meanings. Here, word forms that are similar to each other are mapped onto meanings that are similar to one another, but the specific mapping between the form space and the meaning space is arbitrary.

The fact that even motivated word-meaning associations might become conventionalized should suggest the broadening of the number of potential mappings between sets of words and sets of meanings, and thus their pervasiveness in the language, afforded to mappings between sets of words and meanings that are systematic, but not motivated. Returning to our example of ‘honor’ and related abstract terms for example, I suggested that finding a crossmodal association that could be mapped to these kinds of meanings via the speech channel would be

difficult. However, relying instead on non-motivated associations opens up the number of possibilities for systematic mappings enormously.

Phonaesthemes are one example of a systematic mapping between words and meanings, and have been documented in a number of languages (English: Bolinger, 1980; Indonesian and other Austronesian languages: McCune, 1983; Blust, 1988; Swedish: Abelin, 1999; Japanese: Hamano, 1998; and Ojibwa: Rhodes, 1981). Bergen (2004) suggests that determining the overall proportion of languages that contain phonaestheme clusters is difficult, but that no systematically studied languages have been found to lack phonaesthemes (but, cf. Cuskley, 2013). Consider, for example, the *gl-* phonaestheme cluster in English, which contains a number of words ('glimmer', 'glitter', 'glisten', 'glint', etc.) that have to do with light and vision (Bergen, 2004): we know, based on our previous discussion, that there are in fact well attested crossmodal associations between sound and light (bright = high pitched, for example; Lindauer, 1990). English, however, and other non-tonal languages, do not have the ability to capture this association in a motivated way (whether any tonal languages make use of this motivated association is an open empirical question). By applying a non-motivated systematic mapping however, English is able to capture the similarity between terms denoting light and vision and map that similarity onto a set of similar words beginning with '*gl-*'. However, as I acknowledge above (and, as discussed in Cuskley, 2013) the distinction between phonaesthemes and onomatopoeia is not always clear: '*gl-*' is unlikely to be related

to meanings having to do with light by any motivated process, but instead by historical accident and the process of conventionalisation (Cuskley, 2013). This is not, however, true of all phonaestheme clusters, some of which might be motivated. The ‘sn-‘ cluster for example has a number of words (sneeze, snuffle, snot, snarl) having to do with the nose, and these meanings are mapped to a ‘sn-‘ sound that is nasal; in fact, evidence for this specific motivated systematic association has been demonstrated crosslinguistically (Blasi et al., 2014): because the mapping underlying the ‘sn-‘ cluster is likely motivated, the cluster would be incidentally systematic.

Finally, it bears noting that phonaestheme clusters are not systematic in the absolute sense: the ‘gl-‘ onset is disproportionately associated with meanings having to do with light and vision, but not all words that share the onset share similar meanings (e.g. ‘glove’, ‘glaive’, ‘gloat’, etc.). The systematicity of these types of associations can thus best be characterized as being statistical in nature, rather than absolute, and statistical associations between word forms and meanings have been demonstrated in large swaths of the lexicon (e.g. Monaghan et al., 2014).

1.3.4 Non-Motivated Non-Systematic (Arbitrary)

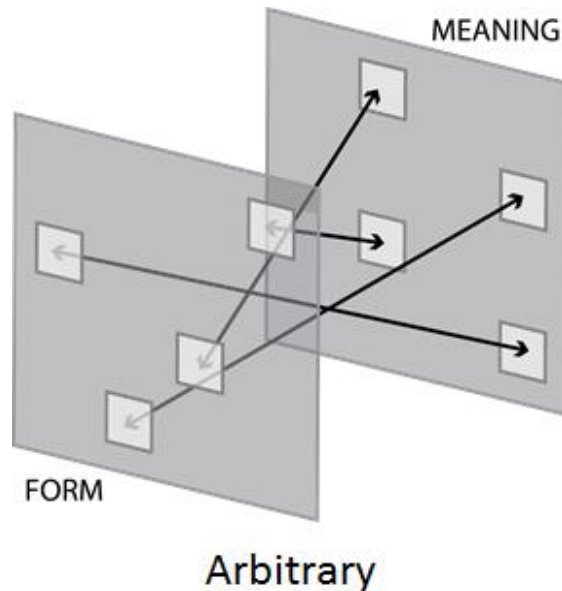


Figure 1.13- A diagrammatic representation of a set of arbitrary mappings between word forms and meanings.

Although non-motivated systematic associations between sets of words and sets of meanings are substantially more flexible than their motivated counterparts, they still theoretically suffer from the issues related to how they constrain the dimensionality of the signal system that I described above: if we assumed that motivated mappings were mandatory where possible then the use of the ‘sn-‘ phonaestheme would limit dimensionality in two ways. First, no word relating to the nose or nasal-oral cavity could begin with any other onset, and second, the ‘sn-‘ onset could not be used for any meanings not related to the nasal-oral cavity.

Relaxation of the motivatedness constraint would allow any onset to be mapped systematically onto meanings related to the nasal-oral cavity, but dimensionality would still be constrained: whatever onset was used for the systematic mapping would be removed as a possible onset for other words.

Arbitrary associations between words and meanings do not have this limitation: the ability to map words to meanings and the learnability of those mappings is not enhanced by either systematicity or motivatedness, but neither is it limited by them. This is ostensibly the explanation for the fact that the lexicon is largely arbitrary: although arbitrariness does not provide any cues for learnability, and thus results in all word-meaning mappings having to be learned in isolation (potentially), motivated and systematic mappings constrain the size of the available signal space and thus might be insufficiently expressive or more difficult to learn in the long run (see below).

The effect that non-arbitrary associations between words and meanings have on learning has been pointed to increasingly in the last decade by researchers suggesting that although motivated and systematic associations account for a relatively small portions of the lexicon, they might be crucial to the process of language learning (Monaghan et al., 2011; Imai & Kita, 2014). Below, I will review the literature suggesting roles for motivatedness and systematicity in language learning, setting the stage for the central questions of this dissertation.

1.4 Motivatedness and Language Learning

In this section I discuss the benefits that motivated associations between words and meanings might provide for naïve language learners. Motivated mappings between words and meanings are central to bootstrapping theories that posit that they scaffold the learning of later non-motivated language (Imai & Kita, 2014; Asano et al., 2015; Perniss & Vigliocco, 2014). I will review this proposal below after summarizing the evidence suggesting that motivated tokens are learned more easily, regardless of any later impact they have on the learnability of non-motivated tokens.

1.4.1 Experimental evidence

There is a wide range of experimental evidence, centered on both artificial language learning and the learning of other motivated word-meaning mappings from unfamiliar languages that supports the notion that motivated associations between words and meanings are easier to learn than are arbitrary ones. In 2012, three papers using very similar methodologies were all published exploring, for the first time, the proposal that the Bouba-Kiki effect might reflect a learning bias, rather than simply a perceptual bias: i.e. that the application of the perceptual bias underlying the Bouba-Kiki effect would enhance learning. Aveyard (2012), Monaghan et al. (2012), and Nielsen & Rendall (2012) all found evidence using their experimental protocols that motivated sound-symbolic associations between consonant plosivity and image

jaggedness were learned more easily than the equivalent counter-motivated configuration of the language. That is, participants were better able to learn that ‘takete’ was the name of a jagged object than they were to learn that it was the name of a curved object, given equivalent training. Nielsen & Rendall (2012) specifically framed their experiment against a common critique in the early sound symbolism literature that sound symbolic biases might be observed precisely because they already exist in some languages but are not motivated (i.e. they are conventionally systematic). This proposal suggest that said biases represent a learned bias, rather than a bonafide learning bias based on the perceptuocognitive organization of human language learners, and based on the results of their experiment, Nielsen & Rendall (2012) suggested that having been learned was unlikely to account for the bias observed experimentally.

Where the three above studies focused on a comparison between motivated and counter-motivated word-meaning associations, other researchers have focused on comparing the learnability of real words from foreign languages that have been judged to be sound-symbolic, vs. words that are arbitrary (but not counter-motivated). Similar to the findings comparing motivated to counter-motivated learning, these studies have shown that motivated crossmodal associations between words and meanings facilitate learning in a number of languages (Asano et al., 2015; Nygaard, Cook, & Namy, 2009). Yoshida (2012) for example, found that non-Japanese speakers learn motivated Japanese words easier than arbitrary ones, and

this effect has also been demonstrated in children (Kantarzis, Imai, & Kita, 2011), and infants as young as 4 months of age (typically assessed through a preferential looking task; Imai et al. (2008); Pena, Mehler, & Nespor, 2011).

1.4.2 Child-directed speech

In addition to findings that learners of languages other than Japanese are able to learn the meaning of Japanese mimetic words at rates higher than would otherwise be expected, there is converging evidence that Japanese mothers preferentially use motivated mimetic terms rather than their non-motivated synonyms when speaking to their infant children (Fernald & Morikawa, 1993; Saji & Imai, 2013). Similar patterns in infant-directed speech are seen in other languages (English, German, and Mandarin Chinese; Grieser & Kuhl, 1988). English speaking mothers, for example, use prosodic cues to exaggerate the pronunciation of words denoting size in way that further enhance motivated associations between size and sound (e.g. ‘huuuuuge’ vs. ‘teeeny’; Perniss & Vigliocco, 2014).

1.4.3 Corpus analysis and crosslinguistic consistency

In addition to findings that mothers preferentially use motivated words when speaking to their infants, further support for the importance of motivatedness specifically for learning can be found in the analysis of motivatedness cross-linguistically. Many languages, including not only Japanese but also some other

Asian languages (Diffloth, 1994; Watson, 2001); indigenous South American languages (Nuckolls, 1999); the majority of sub-Saharan African languages (Childs, 1994); and some aboriginal Australian languages as well (Alpher, 1994; Schultze-Berndt, 2001) have been shown to have large iconic portions of their lexicon. Alone, the presence of motivated associations between words and meanings would not suggest a benefit for learning from those associations, but there has been increasing evidence demonstrating that these sound symbolic pockets of the language are often learned early in development (Thompson et al., 2012). The structure of basic vocabularies, i.e. those that are learned earliest, have also been demonstrated to have shared crossmodal associations (Wichmann et al., 2010), and this finding also extends to some sign languages, where early acquired signs are often the most iconic (Vinson et al., 2008).

1.5 Systematicity and Language Learning

In this section I will review the evidence that systematicity at the level of the lexicon is beneficial for language learning. Although this suggestion is gaining traction rapidly, evidence for it is much rarer in the psycholinguistic literature.

1.5.1 Experimental evidence

There have been very few experimental investigations of the proposal that systematic associations between words and meanings might be beneficial for

learning. Monaghan et al. (2011) stands as the best experimental investigation of this possibility so far: in a series of three experiments Monaghan et al. compared the learnability of artificial lexica that were either systematic (but not motivated) or arbitrary with respect to the mappings between their words and meanings, and found a clear benefit for systematicity in categorization learning. Given a set of words using plosive consonants that is mapped to a set of meanings (e.g. nouns) the regularity of mapping between word and meaning spaces allows participants to learn the category structure exceptionally well. That is, in a systematic language where all words for nouns are made up of plosive letters, experimental participants rarely make the error of assigning a non-plosive word to a noun. The replication and extension of this finding to explore different aspects of systematicity is one of the central contributions of this dissertation, so I will return to these findings multiple times henceforth and with greater precision; at first approximation though, we can count these findings as evidence for a systematicity benefit.

Additionally, some experimental results reported to reflect motivatedness might actually reflect systematicity (Yoshida, 2012). Even the experimental results reported to be traceable to sound-symbolic biases like the Bouba-Kiki effect can be explained partly with reference to systematicity, as the artificial languages used in those experiments often have the characteristic of being incidentally systematic. For example, Maurer et al. (2006) used the set of words ‘k[^]te’, ‘keki’, ‘tite’, and ‘t[^]kiti’ for jagged images.

1.5.2 Computational models

In addition to an experimental exploration of the effect of systematicity on language learning, Monaghan et al. (2011) made use of a series of computational models designed to explore more generally the benefits of systematicity for learning. Broadly, Monaghan et al.'s simulation findings align with those of their experimental participants: systematic associations between word spaces and meaning spaces make categorization easier. These results also align with the findings of an earlier computational model by Gasser (2004), which explored what I have here called incidental systematicity. In Gasser's model, similar signals were mapped onto similar meanings such that the motivated associations between signal spaces and meaning spaces were systematic: we will return to the results of these two simulations in Chapters 2 and 3.

1.5.3 Corpus analysis

Just as corpus analyses have shown an association between age of acquisition and motivatedness, a number of studies have shown that the early acquired part of the lexicon is more systematic than would otherwise be expected (e.g. Monaghan et al., 2014). This finding suggests that the potential benefits of systematicity are leveraged early in language acquisition when they are most beneficial, and is, like

the corpus findings regarding motivatedness, often interpreted as evidence that acquisition of non-arbitrary words bootstraps later learning.

1.6 Motivatedness, systematicity, and language learning

Taking into consideration only the above pressures for learnability and how they would favor the creation of motivated and systematic associations between words and meanings, we would arrive at a model that would suggest that the lexicon would be largely non-arbitrary in nature. Considered mostly simply, the pressure for learnability seems to suggest that arbitrary associations have no advantage and thus that they shouldn't really exist in the lexicon (Figure 1.14).

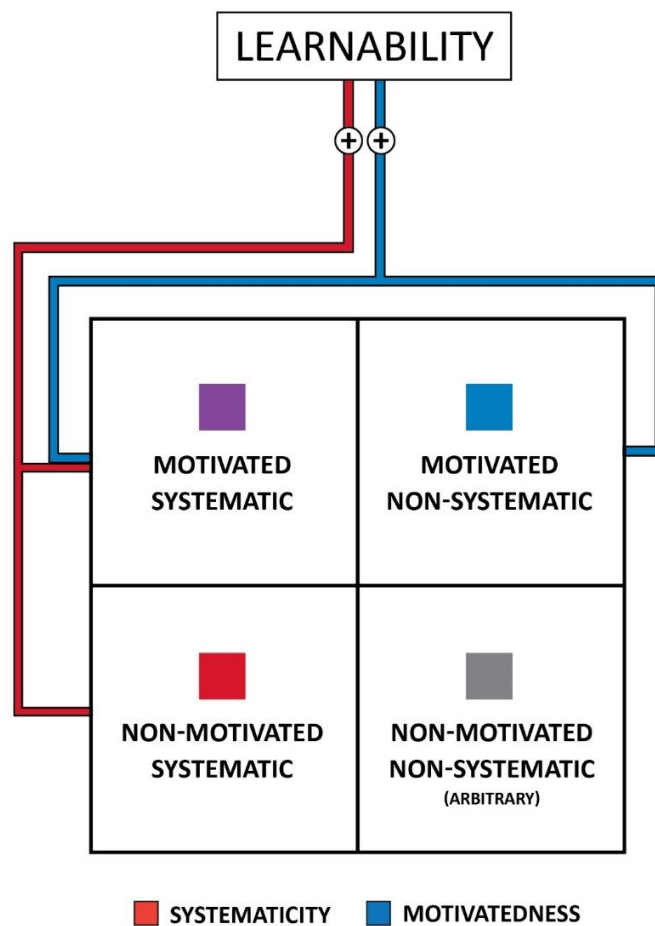


Figure 1.14- A model of the structure of the lexicon based on the pressures of learnability favoring both motivated and systematic associations between words and meanings predicts that the lexicon should be largely non-arbitrary.

This model suggests that the pressure for languages to be learnable interacts with systematicity to produce a positive pressure towards lexica being systematic, containing both motivated systematic and non-motivated systematic tokens. Similarly, because motivated associations are easier to learn, this model suggests

that we should find both motivated systematic and motivated non-systematic word-meaning mappings in the lexicon.

The suggestion of this model that the lexicon should be non-arbitrary, however, is misaligned with the fact that actual lexica are generally reported to be largely arbitrary. In this section I consider the above evidence regarding the benefits and limitations of systematicity and motivatedness to propose a more robust model that accounts for the fact that lexica are largely arbitrary while simultaneously acknowledging the importance of non-arbitrariness for learning. Finding support for this model and determining the relative strength of its components will be the central goal of the remainder of this dissertation.

Non-arbitrary associations between words and meanings are proposed to be selected for by the pressure for languages to be learnable (Imai & Kita, 2014; Monaghan et al., 2011). Thus, the process of cultural transmission should ensure that insofar as these benefits are real we should find them in the structure of the world's languages, and should be able to observe similar phenomena in the lab. However, in addition to a pressure for learnability, languages also have a pressure for expressivity: a language with a single word that applies to all meanings is, as has been pointed out elsewhere (Kirby et al., 2015), perfectly learnable, but also perfectly inexpressive. Figure 1.15, below, represents a more complete model that accounts for these factors and also aligns with the observed structure of the lexicon.

Below, I will tackle the parts of this model individually, pointing towards those claims that require the further experimental support provided in this dissertation.

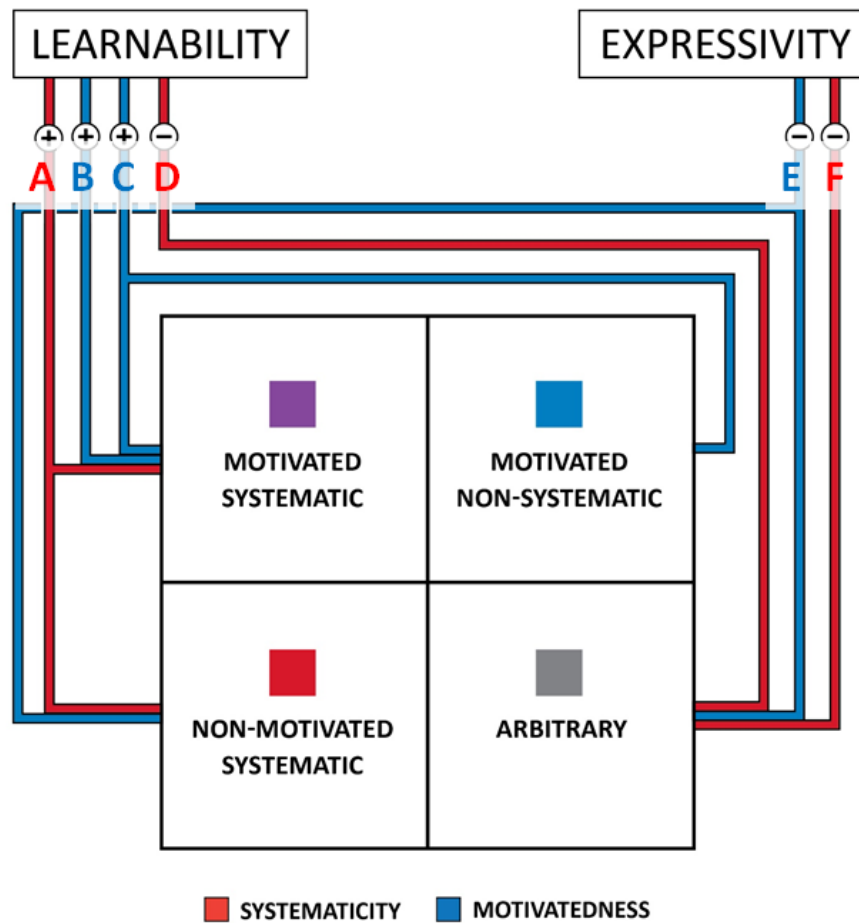


Figure 1.15- A more robust model of the pressures for learnability and expressivity and how they interact to shape the structure of the lexicon. Red lines represent the interaction of systematicity with the pressure from which they originate, and blue lines similarly represent the interaction of motivatedness and their origin. Finally, each line has a notional “valence”, such that lines labeled with a “+” sign reflect outcomes towards their source, and lines labelled with a “-” sign reflect outcomes that bias against their source. For example line A reflects the fact that a positive learnability pressure selects for systematic associations between words and meanings.

Line D, on the other hand, reflects the fact that systematic word-meaning associations can actually make learning more difficult, and thus that the pressure for learnability can select for arbitrary word-meaning mappings.

As acknowledged above, the pressure for languages to be learnable explains the presence of both motivated and systematic associations in the lexicon (Figure 1.15-lines A, B, and C). The evidence for this fact has mostly been covered in the preceding sections of the introduction, but one important question concerning those pressures remains to be addressed: are findings like those of Nielsen & Rendall (2012), which suggest a learning benefit for the Bouba-Kiki effect traceable to systematicity, motivatedness, or both? Many experimental findings that have been traced to motivatedness (Maurer et al., 2006; Nielsen & Rendall, 2011; Monaghan et al., 2012; Ahlner & Zlatev, 2010), for example, could be equally well explained by what I have called incidental systematicity. In Chapter 4 I explore this question experimentally.

In addition to benefitting learnability, both systematicity and motivatedness have limitations and attendant costs. Motivated associations between words and meanings are inherently limited in the kinds of meanings that can be expressed (line E), which should lead to the use of arbitrary words. Additionally, the serial application of motivated associations results in the formation of incidentally

systematic divisions (line B), leading to an increase in the number of motivated systematic word-meaning mappings.

Systematic mappings between words and meanings, because they impinge on the dimensionality of the signal space, can lead to both expressivity and learnability penalties. In the most straightforward case, enforcing absolute systematic divisions in the signal space can result in a hard limit on the number of possible meanings of a given type (line F). In the artificial example presented earlier in this chapter, a language using only cVc trigrams with an average-sized phoneme inventory would only have 3 possible labels for animals, for example (and no possible labels for domestic animals or pets). I do not directly explore this expressivity pressure in this dissertation by way of experiment, and exploring this pressure via corpus study is likely impossible. The fact that no languages seem to suffer from an inability to express an adequate number of meanings suggests that this pressure is very strong.

<u>Level of Systematicity</u>	<u># of Possible Words</u>		<u>Percentage of unavailable mapping space</u>
	<u>Large Phoneme Inventory</u>	<u>Average Phoneme Inventory</u>	
Arbitrary	200000	3100	50
Noun/Verb	50000	773	75
Count/Mass	12500	193	87.5
Animate/Inanimate	3125	48	93.8
Natural/Artificial	781	12	96.9
Animals/Non-Animals	195	3	98.4
Domesticated/Wild	49	0.75	99.2
Pet/Farm Animal	12	0.19	99.6

***Table 1.01-** The effect of systematicity on the size of the available mapping space. At each added level of systematicity, an increasingly large percentage of the mapping space becomes unavailable for use, and the total number of possible words within subcategories becomes smaller. At some level of systematicity, the number of possible words within a category becomes insufficient to express the meanings required for the language.*

In addition to a hard limit on expressivity, systematicity can create practical limits on expressivity. Our theoretical language (with an average phoneme inventory) might need words for 30 different animate objects, which is approximately 1% of the possible words in an arbitrarily structured version of that language, but 66% of the 48 possible words in a systematically structured version of the language that marks the animate/inanimate distinction and all levels above it. Although, strictly speaking, this language would be sufficiently large to express the required meanings, choosing 30 words from that constrained signal space would result in words that were very similar to one another. Assuming that the production or reception of words did not have 100% fidelity (a cognitive limitation of human language learners), the fact that these labels are more similar to one another than they would be in a more arbitrary language might induce a learnability penalty that results in selection for an arbitrary lexicon (line D). Assuming that languages must express a given number of meanings (the number of which varies by language) the effect that systematicity has on reducing the possible mapping space can result either in a language being insufficiently expressive or more difficult to learn. Figure 1.16 below

illustrates this fact, and a number of additional concepts that will become important later.

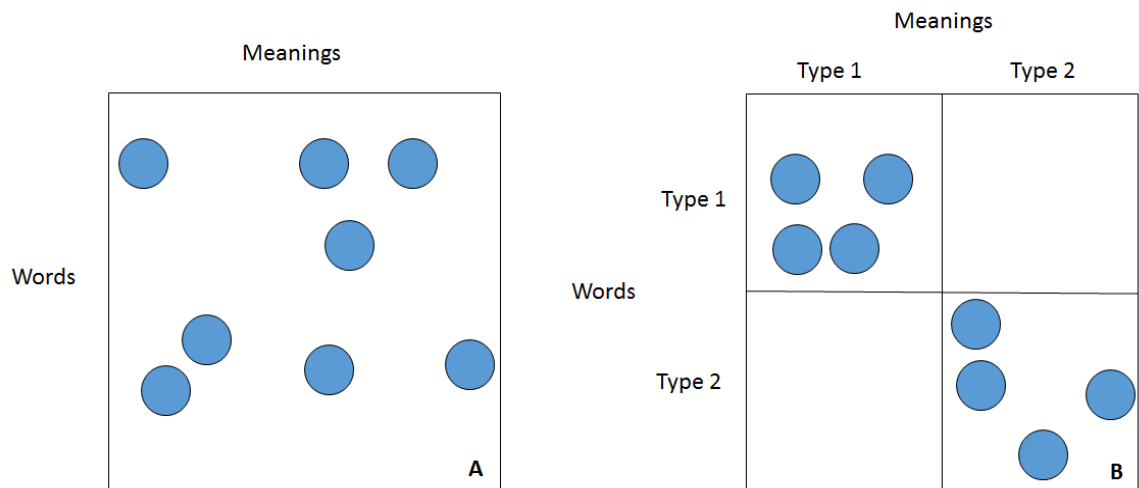


Figure 1.16- *A visualization of the way that systematicity can influence both the size of the available mapping space and the contrastiveness of words within that space. Each circle represents a word-meaning pair in the space with an error term around that word. Where words are closer to each other in the space, they are more easily confused.*

Above, the introduction of a systematic marker in the originally unconstrained signal spaces reduces the size of the overall space for mapping words to meanings. In this case the reduction still allows the language to be expressive (there is space for all of the words), but reducing the size of the mapping space results in words that are, all other things being equal, more similar to each other (and thus potentially more likely to be confused). However, the size of the mapping space does not entirely determine the confusability of the labels: words in the bottom right quadrant are less similar to one another than are words in the upper left quadrant of the signal

space, despite the allowed signal space being the same size. In this dissertation, the similarity of labels to one another will be referred to as contrastiveness: words that are more contrastive should, on average, be learned more easily (and confused less often). In chapters 2, 3 and 5 of this dissertation, I explore this learnability pressure experimentally.

The findings discussed above and the experimental results presented in the coming chapters of this dissertation offer an explanation for the overall structure of the lexicon and why it is predominantly arbitrary, despite learnability pressures generally favoring non-arbitrariness. However, this dissertation also attempts to address the temporal patterning of language acquisition: early acquired words are both more systematic and more motivated than later acquired words, and the question of why this is the case is an important one. One suggestion, discussed below, is that the acquisition of non-arbitrary words early bootstraps the acquisition of the later-acquired arbitrary lexicon.

1.6.1 Bootstrapping

Early research exploring non-arbitrary associations between words and meanings found, repeatedly, that non-arbitrariness aids learning (e.g. Nygaard, Cook, & Namy, 2009) and that in some languages like Japanese, children acquire non-arbitrary words earlier than they do arbitrary ones (Imai et al., 2008). These findings led to the suggestion that the acquisition of the non-arbitrary early

acquired lexicon bootstraps the acquisition of later arbitrary tokens (Imai & Kita, 2014), and to corpus analyses suggesting that the non-arbitrariness of early-acquired portions of the lexicon is not a feature of a very limited set of languages (Monaghan et al., 2014). Some authors have invoked this idea of bootstrapping non-specifically (e.g. Nielsen, 2011) without suggesting any mechanism by which this bootstrapping could occur. At the very least, however, to meet the criteria of bootstrapping we need evidence not only that one event follows another, but also that the learning of the first actually somehow enhances the learnability of subsequent tokens.

The generalized suggestion of bootstrapping - i.e. that learning non-arbitrary tokens enhances the learning of arbitrary ones later, with no invocation of a mechanism will here be referred to as the *simple bootstrapping* hypothesis. Despite the simple bootstrapping hypothesis being artificial and potentially untenable, this dissertation will explore it nonetheless, because of its experimental approachability and what its feasibility might tell us about the feasibility of other bootstrapping hypotheses.

Imai & Kita (2014) have suggested a version of the bootstrapping hypothesis (which they call *the sound symbolic bootstrapping hypothesis*) that I will refer to here as referential bootstrapping. I call Imai & Kita's sound symbolic bootstrapping hypothesis by this different name because I suggest an additional role of sound symbolism for *conceptual bootstrapping* (see below) that is not discussed by Imai &

Kita (2014). The central idea of this hypothesis is that the natural connection between certain words and meanings endows language learners with the ability to establish reference, and that this general referential ability underpins the later ability to map arbitrary words to meanings (Perniss, Thompson, & Vigliocco, 2010). The referential bootstrapping hypothesis thus relies on motivatedness as a mechanism to explain how reference is established (Baldwin, 1993), and further suggests that referential bootstrapping in spoken language is analogous to the enhancement of referential establishment in gestural communication systems (Perlman, Dale, & Lupyan, 2015).

Strangely, no authors have offered up an equivalent bootstrapping hypothesis based on systematicity, rather than motivatedness. Here, I propose that if systematicity also enhances later learnability, it may do so through a process that bootstraps the acquisition or transparency of concepts and categories. I refer to this possibility as *conceptual bootstrapping*, and suggest that systematic associations might make the structure of the underlying categories that they reflect more apparent, or allow for the establishment of categories that are increasingly obscure. An early systematic mapping of some feature of words to, for example, all nouns might be later differentiated to make obvious the differences between count and mass nouns. Similarly, an establishment of an understanding of the difference between nouns and other types of meanings might allow language learners to subsequently learn less systematic or even wholly arbitrary word-meaning

mappings. Further, I suggest that the establishment of incidental motivated systematic associations could underpin the ability of language learners to later learn non-motivated systematic associations - this is a second type of bootstrapping mediated by motivated associations, and the reason that I suggest that Imai & Kita (2014)'s sound symbolic bootstrapping hypothesis is too generally named (because they do not explore this possibility).

Despite raising the possibility of conceptual bootstrapping, I am generally critical of bootstrapping hypotheses because direct empirical support for them is so far lacking. It may be the case, for example, that non-arbitrary associations between words and meanings are in fact easier to learn, but that the subsequent learning of arbitrary associations is underpinned by general cognitive development, rather than bootstrapping. At an early stage of development, non-arbitrary associations might be easier for children to learn due to their limited cognitive ability. With additional time for cognitive development, children might subsequently become increasingly able to learn arbitrary word-meaning mappings. This account would still explain the observed structure of the lexicon and the time course over which its components are acquired, but would not require bootstrapping, which proposes that early learning accounts for an enhancement of later learning, rather than simply occurring subsequently.

1.6.2 Contrastiveness and learnability

In addition to the possibility that the appearance of bootstrapping effects are an illusion caused by cognitive development generally, the transition from early acquired non-arbitrary mappings to later-acquired arbitrary ones might instead be a consequence of the pressures for learnability and expressivity acting on the structure of language. Given this account, early-acquired portions of the lexicon could be non-arbitrary to enhance learning, but these non-arbitrary mappings might create conditions under which the language fails to be adequately expressive or the learning benefit for these associations inverts. In the example discussed above, with 48 possible labels for animate objects, selecting only 3 words that were still relatively distinct from one another would be relatively easy. As children acquire language more completely, however, they will likely require more than 3 labels for animate objects, and additional meaning dimensions will become increasingly salient or necessary. To avoid the possibility that the language reaches either a hard limit on expressivity (insufficient possible words to express all required meanings) or a limit on learnability based on loss of contrast, later-acquired portions of the lexicon might relax their non-arbitrary rules. Looked at over a developmental time course, the resultant mapping between time of acquisition and arbitrariness would seem to support the presence of bootstrapping, but here this structure would be the result of an expressibility pressure, rather than bootstrapping. I explore the difference

between an explanation relying on bootstrapping and one accounted for only by the pressure for learnability directly in Chapter 5.

1.7 Thesis Outline

The experiments presented in this thesis will focus primarily on the question above: how much can the interaction of pressures for learnability and expressivity tell us about the structure of language. The influence that these pressures have on the structure of language will also be evaluated with respect to bootstrapping hypotheses to assess their feasibility or necessity.

1.7.1 Overview and contribution statements

The body of this thesis has been written specifically for this purpose, other than the text of Chapter 4, which has been submitted to the *Journal of Experimental Psychology: Learning, Memory, and Cognition*. The body of that chapter appears in the form that it will appear in publication, other than some minor editing to adhere to the format of the remainder of the dissertation. However, as that chapter is based on a publication, the writing of that chapter was shared more evenly between myself and its other authors, who were material not only in the writing but also the statistical analyses presented therein.

Additionally, the model presented in Chapter 3 of this dissertation was created collaboratively with Dieuwke Hupkes, a visiting master's student who was

responsible for the coding of the model, under the supervision of myself and Dr. Kenny Smith.

1.7.2 Chapter 2- Systematicity and language learning

In Chapter 2, I present the results of a series of experiments exploring the benefits of systematicity for the learning of artificial lexica. In Experiment 1 I straightforwardly extend the design of Monaghan et al. (2011) using a new experimental paradigm that allows for an easier exploration of the learnability penalty for systematic languages due to the confusability of similar words. In Experiment 2, using maximally contrastive but still systematic artificial lexica, I demonstrate that systematic mappings that maintain contrastiveness can allow for a learnability benefit that does not collapse the signal space and result in increased confusability of within-class words. Finally, in Experiment 3, I attempt to directly manipulate contrastiveness and the degree to which introducing a systematic mapping constrains signal space dimensionality, and find further support for the general conclusion that systematicity can enhance learnability in terms of categorization but not always individuation. In that experiment I also demonstrate that the relationship between contrastiveness and signal space saturation is not necessarily entirely linear, and thus that new metrics are required to compare these factors adequately.

1.7.3 Chapter 3- Phonological dispersion, systematicity, and language learning

In Chapter 3 I present the results of a further extension of Monaghan et al.'s experimental paradigm, and also a recreation of their computational model that explores a novel form of systematic mapping between words and meanings that is systematic but not based on phoneme feature similarity. I find that although this novel type of systematic association between words and meanings results in massively different learning by the computational model, my experimental participants learn systematic associations similarly regardless of what dimension they are structured along.

1.7.4 Chapter 4- Motivated vs. Conventional systematicity

In Chapter 4 I present an experiment designed to test the difference in learnability between motivated systematic artificial lexica and lexica that are systematic, but not motivated. I find that overall there is no difference in learnability between the two types of lexica, although participants in the motivated condition of the experiment have an advantage on early trials where they are able to respond based on perceptual bias. Additionally, I find that the presence of motivated associations between words and meanings can interfere with the learnability of non-motivated systematic associations along a second dimension. I interpret these results to suggest that the potential of motivatedness to interfere with expectations regarding

motivatedness more generally might account for the fact that expressives in natural languages are often markedly different in their phonotactics: by isolating motivated portions of the lexicon from the rest of the lexicon, any negative influence that they have on learnability can be minimized.

1.7.5 Chapter 5- Growing Lexicon experiment

In Chapter 5 I present an experiment that allows for an exploration of the time course of learning to test the bootstrapping hypothesis and also how the learnability of words varies as a function of their likelihood of being confused with other words of the same type. The results of this experiment suggest that the simple bootstrapping hypothesis does not seem to account for the observed learnability of arbitrary tokens subsequent to the learning of motivated systematic ones, but rather that contrastiveness and confusability alone account for this finding.

1.7.6 Conclusions

I conclude the dissertation by briefly rehearsing findings of the experiments presented here and situating them in the fields of linguistics and psychology more generally, suggesting that ultimately the fundamental pressures of learnability and expressivity interact to shape the structure of language, and that bootstrapping hypotheses are not required to account for the fact that arbitrary word-meaning mappings are learned later than non-arbitrary ones and account for a much more

substantial portion of natural lexica. I conclude, however, by suggesting that both perceptual and conceptual bootstrapping likely account for some of the observed properties of language and acquisition trajectory of its learners, suggesting experimental protocols that might allow for future exploration of this possibility and establishment of how it interacts with expressivity and learnability pressures.

Chapter 2

Systematicity and Learning I

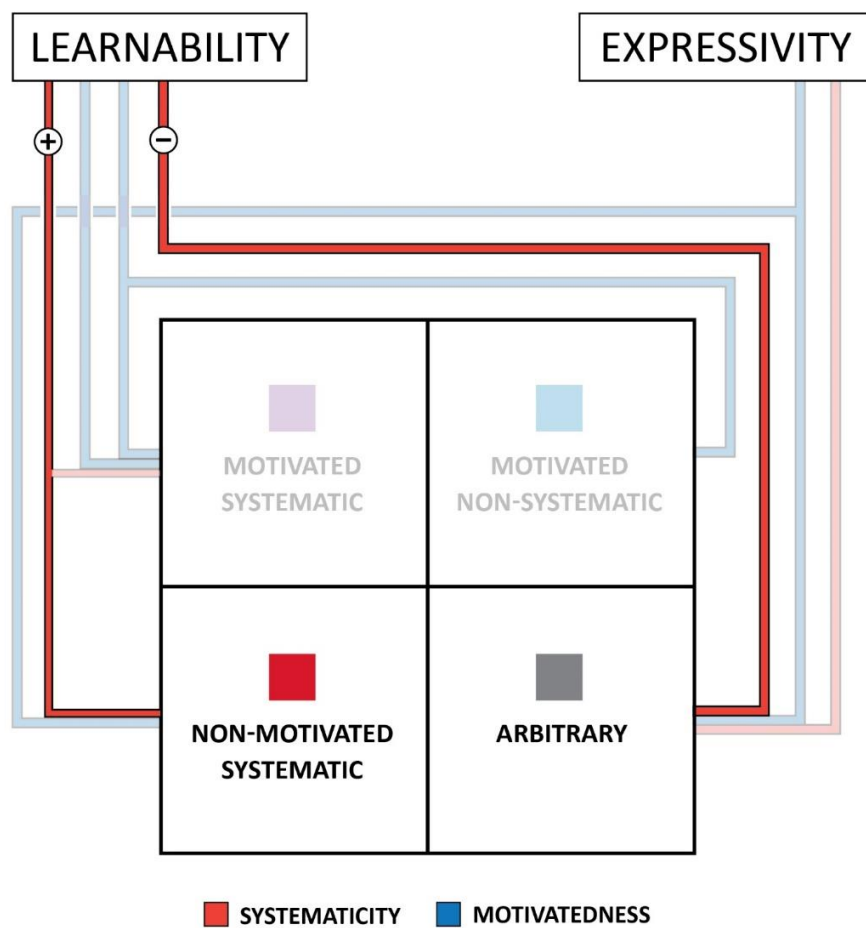
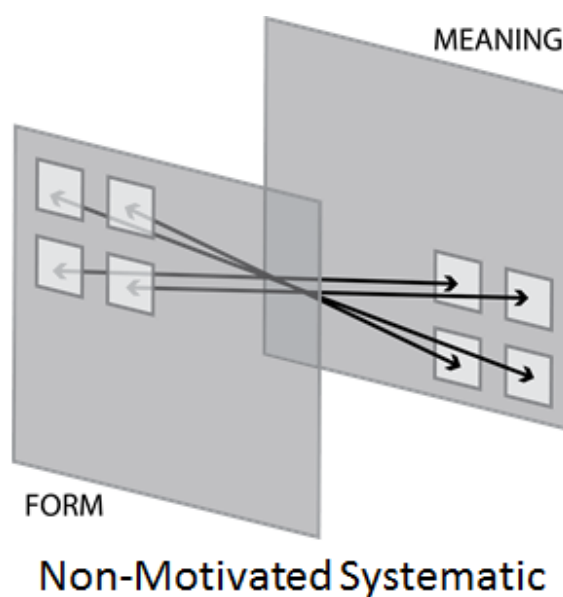


Figure 2.01- Chapter 2 compares the effects of learnability pressure towards systematicity to the effect of the learnability pressure towards arbitrariness by comparing the learnability of systematic (but non-motivated) artificial languages to arbitrary ones.

This chapter focuses on an exploration of the effect that various types of systematic, but non-motivated associations between words and meanings have on learnability. As discussed in Chapter 1, in the past decade there has been an increasing interest in the effect that systematic mappings between words and meanings have on learning. To rehearse, systematicity refers to any mapping of words to meanings such that a feature shared by a set of similar words is reliably associated with a feature shared by a set of similar meanings.



***Figure 2.02-** A diagrammatic representation of a non-motivated but systematic association between word forms and meanings. Here, word forms that are similar to each other are mapped onto meanings that are similar to one another, but the specific mapping between the form space and the meaning space is arbitrary.*

In this chapter, I explore systematic associations between words and meanings that are non-motivated. For example, there is probably nothing that makes ‘gl-’ a

particularly good segment for words having to do with light, but nonetheless there is a cluster of words beginning in ‘gl-’ that share similar meanings (‘glimmer’, ‘glare’, ‘glint’, ‘glow, etc.). Specifically, this chapter explores the learnability pressures for systematic associations: both the learnability benefit for systematic associations (Figure 2.01 A) and the learnability/practical expressivity cost associated with those same types of associations (Figure 2.01 B).

To this end, I present the results of three experiments exploring the learnability of non-motivated systematic artificial languages compared to arbitrary (non-motivated, non-systematic) languages. The evidence provided in this chapter focuses on a fundamental split of learning into two types. *Individuation* refers to what we typically thinking of as word learning – assigning the correct word to a given meaning: for example, recognizing ‘glare’ as the correct word to describe a strong and bright light. *Categorisation*, on the other hand, refers to assigning the incorrect word to a meaning where that incorrect word is still of the same type as the correct word – for example, to accept the word ‘glare’ as referring to a sparkly reflected light (‘glitter’). The results of these experiments suggest the presence of a general learnability benefit for systematic lexica (Figure 2.01-A) based primarily on a benefit for categorisation learning, and a learnability penalty contingent on the degree to which systematicity impinges on practical expressivity/learnability (Figure 2.01-B).

2.1 Background and Rationale

As we saw in Chapter 1, systematic associations between words and meanings can divide the lexicon in such a way that they impose limits on expressivity. The first of these types of limits, which I referred to as a hard limit on expressivity, occurs when the number of words allowed by a systematic division is smaller than the number of individual meanings of that systematically marked type that need to be expressed. In our artificial language from Chapter 1 with an average phoneme inventory (22.7 consonants and 6 vowels) used only trigrams, there would be just over 3000 possible words for that language. If this artificial language systematically marked words for animals as beginning with the segment ‘pI’, there would be 23 possible words for animals (pIg, pIp, pIf, pIn, etc.; assuming no phonotactic constraints disallowed certain combinations). If this language required names for 30 animals, this systematic mapping would impose a hard limit on expressivity: it simply would not be possible given these constraints to have unique names for each animal. If this language required names for 18 animals, however, expressivity would theoretically be unhampered - it would be possible to assign names to each of the 18 animals. These 18 words, however, would all be very similar to one another, and thus might be difficult to keep separate, especially given their similar meanings.

This tension between a learnability benefit for systematic word-meaning mappings and a learnability penalty induced by a loss of contrastiveness is the

central issue of this chapter, and has been explored previously both experimentally (Monaghan et al., 2011) and through computational modelling (Gasser, 2004; Monaghan et al., 2011). Both of these approaches to the tension between learnability benefits and penalties suggest that the strength of the two pressures is contingent on how much systematicity constrains the signal space. Systematic mappings between words and meanings aid the process of categorisation: in the above example, upon hearing the word ‘pIk’ a learner familiar with the language would know that it referred to an animal, even if they had never heard the word before. However, if asked to feed to ‘pIm’ a user of this language might find themselves confused: is ‘pIm’ the right word for a horse, a cow, a cat, a chicken, a rooster, or a crocodile?

In 2011, Monaghan and colleagues attempted to explore this tension directly. In a series of experiments, Monaghan et al. (2011) compared the ability of human participants to learn languages that were either entirely systematic or entirely arbitrary. Systematic language learners in Monaghan et al. (2011)’s first experiment learned names for a set of nouns that were made up of a small set of phonemes, and words for a set of verbs that were made up of a second small set of phonemes, with no overlap between the two. In this experiment, Monaghan et al. found that learners of systematic languages had an overall advantage for categorisation over learners of arbitrary languages. In addition to a benefit for categorisation, learners of systematic languages showed an early individuation benefit, but arbitrary language learners eventually matched the overall performance of systematic language

learners. Monaghan et al. (2011) also included a number of computational models of the same kind of language learning that matched the general patterns of learning observed in their human learners, but in this chapter I will focus on their experimental results (although I return to the model results in Chapter 3).

The findings of Monaghan et al. (2011)’s first experiment are limited in their ability to separate the learnability benefit of systematicity from the learnability penalty based on loss of contrastiveness because they explore on a single signal space. Specifically, the signal space used by Monaghan contains only 16 possible words, of which 12 were used; thus, all words tested were very similar to one another, and this might have inflated the learnability penalty.

In 2004, Michael Gasser created a computational model of language acquisition that allowed for an exploration of the difference in learnability between systematic and arbitrary languages where the size of the signal space and the vocabulary also varied. In a signal space with 1000 possible “words”, Gasser (2004) found that when vocabulary was small (15 word-meaning mappings) systematic associations between words and meanings were easier to learn than arbitrary ones, but with a larger vocabulary size (100 words) arbitrary languages were easier to learn than systematic ones. However, when Gasser increased the size of the possible signal space to 10,000 possible words, he found that the large vocabulary condition still favored systematicity over arbitrariness for learning.

2.2 Chapter Outline

In Experiment 1, I present the results of an extension of an experimental paradigm used by Monaghan et al. (2011) where experimental participants learn a language that is either systematic or arbitrary, and find, supporting Monaghan et al. (2011) that systematic languages are easier to learn, but that this learning benefit is based primarily on those languages allowing for categorisation. That is, systematic languages are easier to categorise, but because their words are more similar to one another, more difficult individuate than arbitrary languages.

The words learned in Experiment 1 (as well as in Monaghan et al., 2011) are very similar to one another, being chosen from a relatively small signal space. As we saw in the introduction, the similarity of these words to each other based on a systematic mapping between words and meanings might have a negative impact on learnability. Thus, in Experiment 2 I present the results of a further extension of the experimental methodology used in Experiment 1 that tests the learnability of languages that are systematic, but chosen from a much larger potential signal space such that they are less similar to one another. I find, using these more contrastive stimuli, that systematic languages retain an advantage for categorisation without incurring the same penalty to individuation.

Monaghan et al.'s 2011 paper includes two additional experiments exploring different types of non-motivated systematic mappings between words and meanings.

In Experiment 3 of their paper, they introduce an experiment where systematic marking is somewhat relaxed such that labels within categories can be more contrastive while maintaining a strictly systematic construction. In my Experiment 3, I attempt to further explore this possibility by constructing an experimental paradigm to manipulate the contrastiveness of labels to one another and delineate the degree to which overall contrastiveness influences learnability, although the metric by which I calculate contrastiveness makes interpreting the results of that experiment difficult.

2.3 Experiment 1

Experiment 1 focuses primarily on a replication of the experimental results of Monaghan et al. (2011), but using a slightly different experimental methodology. Where Monaghan et al. used a forced choice task for testing where participants were presented with a single word and tasked with choosing the correct meanings for that word from all possibilities; Experiment 1 here uses a signal detection paradigm to evaluate learning. The use of a signal detection paradigm, where on each trial participants are presented with a single word and a single meaning and asked to either accept or reject the pairing as correct, allows for a number of tests comparing individuation learning to categorisation learning that are not as straightforward using an alternative forced choice task. In addition to the use of a new experimental methodology (signal detection) Experiment 1 differs from Monaghan et al. (2011) in

the specific stimuli used- the differences between my stimuli and Monaghan et al.'s will be described below in the relevant sections. Finally, Monaghan et al. used a static selection of words from their available signal space: although there were 8 possible words of each type, all participants were taught the same subset of 6 of those words. In Experiment 1 I relax that control, resulting in slightly different languages for each experimental participants: for each participant 8 words of each type were chosen from 64 possible words. Because of this, the similarity of each participants words to each other varies slightly, allowing for an exploration of the effect that this subtle difference in contrastiveness has on learnability.

2.3.1 Methods

Participants

Participants were 26 students (11 female) recruited from the general population of the University of Edinburgh, and were compensated 2.00 GBP for the 15 minutes required to complete the task. All participants were monolingual English speakers between 17 and 31 years of age. Ethical approval was obtained from the University of Edinburgh in line with British Psychological Society (BPS) guidelines, and informed consent was obtained from all experimental participants.

Participants were assigned randomly to each of 3 experimental conditions.

Conditions 1a (n=7) and 1b (n=7) were counterbalanced systematic language

conditions. We found no differences between participants in these subconditions (i.e. it did not matter whether animals were paired with plosive or sonorant words), so those subconditions were collapsed for further analysis. The remaining participants (n=12) were assigned to Condition 2 (arbitrary language).

Experimental Design

Label Stimuli

Monaghan et al. (2011) created two types of words that differed in both their consonant and vowel composition. From a set of four consonants (/f/, /ʒ/, /g/, and /k/) and four vowels (/i/, /ɪ/, /u:/, and /a:/) two types of words were created, each with eight possible words (six of which were used: see Table 2.01). I created the new label stimuli for Experiment 1 using similar constraints, although the labels created were trisyllables (in cVcVcV configuration) rather than trigrams, and created using a slightly larger and different set of phonemes.

Words of the first type were constructed from the obstruent consonants /t/, /k/, and /p/ in combination with the vowels /i/ and /e/ while words in the second class were constructed from the sonorant consonants /m/, /n/, and /l/ and the “rounded” vowels /o/ and /u/.

For each of the possible consonant phoneme positions, two of the available phonemes were chosen as possibilities for label construction while for each vowel the

two possibilities remained the same (see Table 2.01), resulting in a total space with 64 possible words of each type. For each experimental participant, a set of 8 words of each type was chosen from this total space (Table 2.02), giving over 4 billion possible combinations of words of each type. By contrast, Monaghan et al.'s signal space had 8 possible words, of which 6 were used, which gives 28 possible combinations of words of each type (of which only 1 combination was used).

	Consonant 1	Vowel 1	Consonant 2	Vowel 2	Consonant 3	Vowel 3	Sample Labels
Type 1- Plosive + Non-Rounded	/t/	/i/	/p/	/i/	/t/	/e/	'tipite'
	/k/	/e/	/k/	/i/	/p/	/i/	'kekipi'
Type 2- Sonorant + Rounded	/m/	/o/	/n/	/o/	/l/	/u/	'monolu'
	/l/	/u/	/l/	/u/	/n/	/o/	'luluno'

Table 2.01- Word construction procedure for Experiment 1.

Monaghan		Experiment 1	
Type 1	Type 2	Type 1	Type 2
Systematic	fɪʒ	ga:k	kekete lomolo
	flɪʒ	gu:k	kekiti lomumu
	ʒɪf	ka:g	kepipe lumomo
	ʒɪf	ku:g	kepipi molulo
	ʒɪʒ	ka:k	kikite molumo
	flɪf	gu:g	kipete momomo
			tekete momulu
Arbitrary	fɪʒ	flɪʒ	kekete mulomo
	ʒɪf	ʒɪf	kepipi kikite
	flɪf	ʒɪʒ	kekiti tekipe
	ku:g	ga:k	kipete kepipe
	ka:k	gu:k	tekete kekete
	gu:g	ka:g	molulo lomolo
			molumo momomo
			mulomo lumomo
			lomumu momulu

Table 2.02- A comparison of the word stimuli used in Monaghan et al. (2011) (left) and Experiment 1 (right).

Acoustic stimuli for each of the words was created using Apple talk with the female voice Victoria. Because Apple talk does not use phonetic symbols, the actual pronunciation of the words shown here is somewhat inexact, although I did my best to ensure that the phoneme representations were accurate. At the very least, the actual words produced by apple talk were discriminable such that the phonemes for systematic language learners were systematic (i.e. the pronunciation of /u/ was not always precise, but it was never produced similarly to /i/ or /e/).

The similarity of the set of words for each participant was calculated using Hamming Distance, which is a measure of the difference between two strings of equal length based on the number of positions where the phonemes are different. Thus, at each consonant or vowel position, each label was compared to every other label and a score between 0 and 6 was calculated, giving the distance from that label. With this information, I calculated an average contrastiveness score for each participant and included in the data for analysis.

Experiment 1 used the sonorant/plosive and rounded/unrounded dichotomy typically found in sound symbolism research relating to the Bouba-Kiki effect (cf. Maurer et al., 2006; Nielsen & Rendall, 2011, 2012) to allow for the extension of the experimental paradigm to exploration of the effects of motivatedness on learnability (see Chapter 4).

Image Stimuli

Image stimuli were also split into two categories: animals and vehicles, and were taken from a variety of online sources using Google Image search; images were placed on a white background, then standardized for size and resolution. The use of two categories of nouns differs from Monaghan et al., who used images depicting actions and objects from the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997). Thus, in addition to having a difference in the contrastiveness of my labels, relative to Monaghan et al., it is possible that there is also a difference in the

contrastiveness of my meanings, although quantifying such a difference would be difficult.

Experimental Design

For each participant, eight of the twelve images and labels of each type were randomly selected as stimuli. These labels and images were then paired together based on the experimental condition. This part of the experimental set-up differed from Monaghan et al. only in that participants in Monaghan et al. learned 12 word-meaning pairs, where participants in Experiment 1 learned 16.

In the **systematic** condition of the experiment, all images of one category were paired with all labels of one category, with the second category of images paired with the second category of labels. For example, all animals could be paired with plosive words and all vehicles with sonorant ones, or the opposite assignment could be applied; which of these pairings was used was counterbalanced across participants.

In the **arbitrary** condition of the experiment, half of the images from each category were paired with half of the labels of each category; thus, half of the animals were given plosive labels, and half were given sonorant labels, with the same being true for the vehicles.

The experiment was programmed and conducted using Livecode v 5.0.2.

Procedure

Familiarisation

Prior to training, participants were familiarised with all of the words that they would subsequently learn the meaning for. For familiarisation, each word was played to the participant via headphones twice, with a 1 second delay between each presentation. The order that the words were presented in was randomized, and participants were given two rounds of familiarisation (for a total of four exposures to each word).

Training

The training portion of the experiment involved the sequential presentation of all of the paired labels and images. On each training exposure, the participant was shown an image in the center of the screen. 750 milliseconds later, the word for that image was played to them via headphones, and then, after a 1 second delay played a second time. After the second presentation of the word, the image remained on screen for one second before progressing to the next trial. Each association was presented twice in randomized order in each of two training blocks.

Testing

Experiment 1 used a signal detection paradigm to measure the ability of participants to learn the associations that they were taught during training. On each test trial participants were presented with a single image in the center of the screen concurrently with the presentation of a single auditory stimulus via headphones. Participants were tasked with responding either “yes” (by pressing the “z” key) or “no” (by pressing “/”) to indicate whether they had previously seen the specific pairing of image and label that they were presented with.

Trials in the test phase of the experiment were split into three types: targets, in-class distractors, and out-of-class distractors. Target trials were those in which the presented image-label pair was one that had been seen during training. In-class distractor trials involved presented images being paired with labels that were different from the one that they had been trained with, but were of the same type (thus, if the word for a given image was made using plosive consonants, an in-class distractor trials would pair that image with another word containing plosive consonants). Finally, out-of-class distractor trials paired images with incorrect words that were of the opposite type. There were a total of 64 test trials for each experimental participant (16 Target, 16 In-class, 32 Out-of-class).

Data Analysis

Participant responses were scored according to a signal detection paradigm; on target trials “yes” responses were scored as hits, with “no” responses as misses, while

on distractor trials of both types “yes” responses were scored as false alarms with “no” responses scored as correct rejections. This type of scoring allowed for the calculation of a d' value for each of the participants, which is a measure of the ability of participants to discriminate between alternatives that effectively controls for experiments where there are many more distractor trials than there are targets. Overall comparison of participants in the two conditions (systematic vs. arbitrary) was compared using a two-sample t-test.

A repeated-measures analysis of variance of response correctness was also conducted with experimental condition as a between-subjects factor and trial type (In-class, Out-of-class, or target) as a within-subjects factor. This analysis allowed me to break down participant responses based on categorisation and individuation. Performance on target trials is a straightforward way to gauge individuation learning, but the difference between performance on target trials and in-class-distractor trials is actually the most relevant comparison. A learner who individuates perfectly will accept all target trials and reject all in-class-distractor trials, while one who has learned only the category structure (e.g. that ‘keketi’ is a word for a vehicle) will accept all target trials *and* all in-class-distractor trials. Performance on out-of-class distractor trials can also be used to gauge the ability of learners to categorise- those who have learned the category structure should be able to easily reject all out-of-class distractors. In terms of our learnability pressures, the learnability benefit suggested for systematicity should lead to systematic language

learners performing better on target trials, out-of-class distractor trials, or both. The learnability penalty suggested for systematicity however should lead to systematic language learners performing significantly worse on target trials, in-class-distractor trials, or both.

Finally, in addition to comparisons of d' and the repeated measures ANOVA, a simple linear regression was conducted for each experiment condition to test the correlation between learnability and the average contrastiveness of a participant's language. I predicted that overall, more contrastive languages would be easier to learn than less contrastive ones (i.e. there would be a positive correlation between correctness and contrastiveness).

2.3.2 Results

Signal Detection

Participants in the systematic and arbitrary language learning conditions both performed at rates above chance. Participants in the systematic condition of the experiment had an average d' score of 1.49 (SD= 0.64) while participants in the arbitrary language condition had an average d' score of 0.42 (SD= 0.39).

Participants in both conditions performed significantly better than chance (systematic: $t(13)=8.75, p<0.001$, arbitrary: $t(11)= 3.66, p=0.0037$), but participants

in the systematic condition performed better than those in the arbitrary condition ($t(24) = 5.06, p < 0.001$; Figure 2.03).

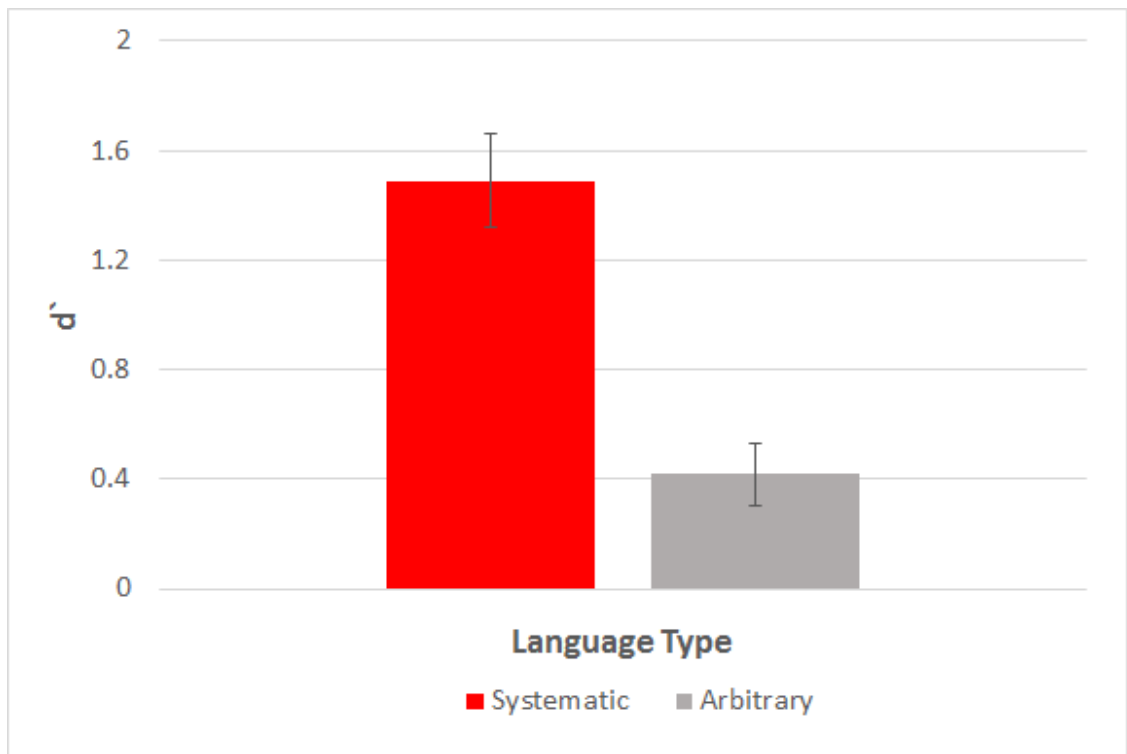


Figure 2.03- *d'* performance by participants in the systematic and arbitrary conditions of Experiment 1 scored by their ability to identify pairs of objects and labels that they had previously learned in the training phase of the experiment. Performance in both conditions was significantly better than chance (both $ps < 0.01$) and participants in the systematic language condition performed significantly better than those in the arbitrary language condition ($p < 0.001$).

Repeated Measures Analysis of Variance

The repeated measured analysis of variance revealed a significant main effect of condition: participants in the systematic condition ($M = 0.67, SE = 0.021$) performed significantly better than participants in the arbitrary language condition ($M = 0.57,$

SE= 0.023; $F(1,77)=9.95$, $p=0.0043$. There was also a significant main effect of trial type ($F(2,77)= 42.13$, $p<0.001$): post-hoc comparison using the Tukey-Kramer Multiple comparison test showed that participants performed significantly worse on in-class-distractor trials ($M=0.42$, $SE= 0.029$) than on either Target ($M= 0.69$, $SE=0.029$) or out-of-class distractor trials ($M= 0.75$, $SE=0.021$).

In addition to these main effects there was a significant interaction between experimental condition and trial type ($F(2,77)= 35.65$, $p<0.001$; Figure 2.04). Post hoc analysis of this interaction showed that participants in the arbitrary language learning condition did not perform significantly differently on the three trial types (Target: $M= 0.59$, $SE=0.043$; In-Class Distractor: $M= 0.55$, $SE= 0.043$, Out-of-Class Distractor: $M= 0.57$, $SE=0.031$; $F(2,732)=0.23$, $p=0.79$). Participants who learned systematic languages however performed significantly differently depending on trial type ($F(2,854)= 92.35$, $p<0.001$): Systematic language learners performed best on out-of-class distractor trials ($M=0.94$, $SE=0.028$), second best on target trials ($M=0.79$, $SE=0.04$) and worst on in-class distractor trials ($M=0.28$, $SE=0.04$) and all of these differences were significant according to the Tukey-Kramer Multiple comparison test.



Figure 2.04- *Effect of the interaction of experimental condition and trial type on the proportion of correct responses. Participants in the arbitrary language condition performed equally well regardless of the type of experimental trial, while systematic language learners performed best on out-of-class distractor trials and worst on in-class-distractor trials.*

Additionally, I compared the performance on each trial type between the two conditions using further rmANOVAs. I found that on target trials, participants in the systematic condition performed better than those in the arbitrary condition (Systematic: $M = 0.79$, $SE = 0.039$; Arbitrary: $M = 0.59$, $SE = 0.042$; $F(1,25) = 11.59$, $p = 0.0023$). Similarly, participants in the systematic condition performed significantly better than those in the arbitrary condition on out-of-class distractor trials (Systematic: $M = 0.94$, $SE = 0.028$; Arbitrary: $M = 0.57$, $SE = 0.031$; $F(1,25) =$

76.95, $p < 0.001$). However, on in-class distractor trials, participants in the arbitrary condition performed significantly better than systematic language learners (Arbitrary: $M = 0.55$, $SE = 0.046$; Systematic: $M = 0.28$, $SE = 0.042$; $F(1,25) = 18.77$, $p < 0.001$).

The fact that performance on in-class distractor trials was significantly below chance for systematic language learners ($M = 0.28$; $t(13) = 4.89$, $p < 0.001$) prompted a final test comparing the inverse of performance on in-class distractor trials to performance on target trials to determine if performance on the two trial types could be explained entirely by categorisation learning. A two sample t-test showed that for systematic language learners correctness on target trials was not significantly different from the inverse of correctness on in-class distractor trials ($t(26) = 1.19$, $p = 0.243$).

Finally, a linear regression revealed that a moderate positive correlation between average contrastiveness and d' ($r = 0.406$, $p = 0.039$; See Figure 2.05), although this was not true for either systematic language learners ($r = 0.23$; $p = 0.43$) or arbitrary language learners ($r = -0.17$; $p = 0.60$) when analysed alone.

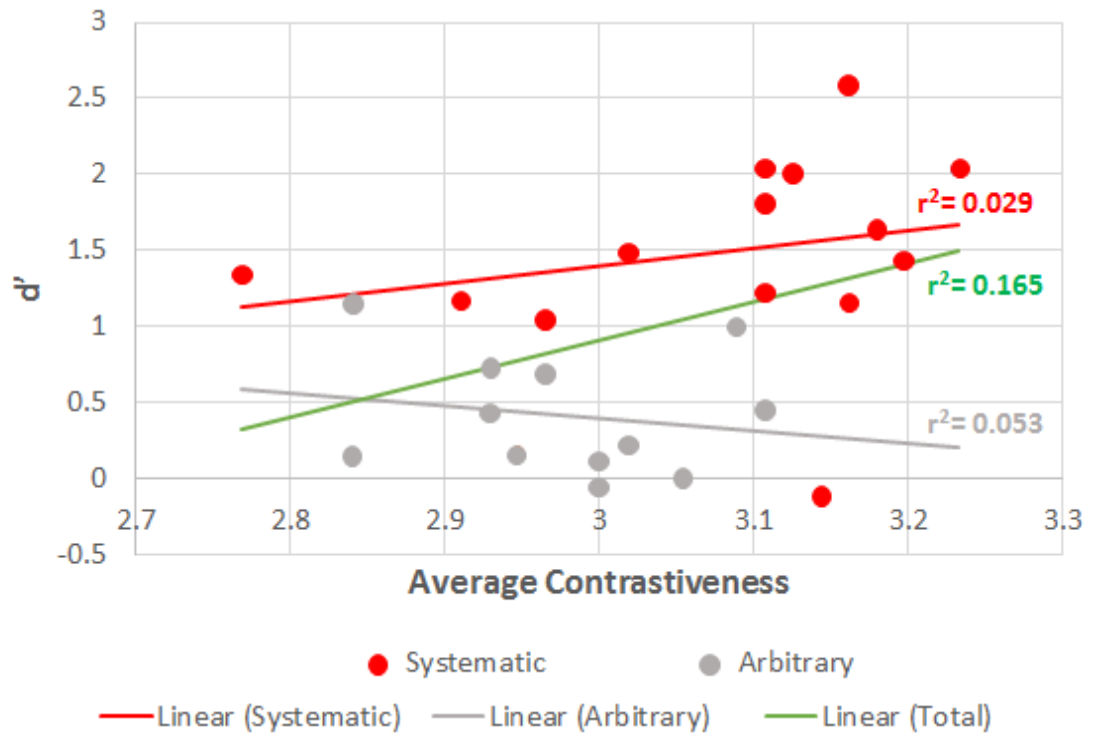


Figure 2.05- Linear regression of d' as a function of contrastiveness. Data from systematic language learners (and their correlation) is plotted in orange, while data from arbitrary language learners is plotted in blue. The green regression line represents the linear fit to the total data set and shows a moderate correlation between the two, with contrastiveness accounting for 16.5% of the variance in d' scores ($p=0.039$).

2.3.3 Discussion

My prediction that systematic language learners would perform significantly better when the data was analysed in terms of categorisation was supported by the data from Experiment 1. Systematic language learners performed significantly better on out-of-class distractor trials but suffered a penalty on in-class distractor trials

relative to arbitrary language learners. In support of a benefit of systematicity for individuation, I found that participants who learned systematic languages performed better on target trials than did arbitrary language learners.

I evaluated the interaction between a benefit for categorisation and a benefit for individuation by comparing performance on target trials and in-class distractor trials for systematic language learners. This analysis showed that systematic language learners accepted target and in-class distractor trials at equal rates, which could be interpreted to suggest that they were able to learn only category information and not able to individuate at all. Subtracting the proportion of accepted in-class distractors (0.72) from the proportion of accepted targets (0.79) suggest that, at best, systematic language learners would be truly individuating correctly on only 7% of trials, although this difference is not significant.

The correlation of contrastiveness to correctness was significant and in the direction that I predicted, although the overall variance in contrastiveness between experimental participants was fairly low. Thus, to better understand this issue an experimental methodology that allows for individual comparisons on a trial-by-trial basis would be required (see Chapter 5).

A direct comparison between the results of Experiment 1 and the results of Monaghan et al.'s Experiment 1 is difficult for a number of reasons, but their general conclusions seem to be fairly well supported by my data. First, and most obviously,

I used a different experimental protocol (Signal Detection vs. AFC) which produced different data for analysis, although the ability to separate individuation and categorisation is easier given my data structure. Second, Monaghan et al. used multiple rounds of training and testing in their experimental protocol, whereas I had only a single round of training and testing: thus the most relevant comparison between our results compares Experiment 1 presented here with the first block of Monaghan et al.'s Experiment 1 (although in their case, participants were given twice as much training).

The graphs below (Figures 2.06 and 2.07) show a comparison of my results and Monaghan's split into categorisation and individuation. For categorisation, I calculated a value based on the average of the number of accepted pairs that were either targets or in-class distractors for each experimental condition. For individuation, I use the proportion correct for my experimental participants across all three trial types. Because these data were obtained using different experimental designs, I do not include any statistical comparisons here, as they wouldn't be appropriate, instead including the comparison only for illustration purposes.

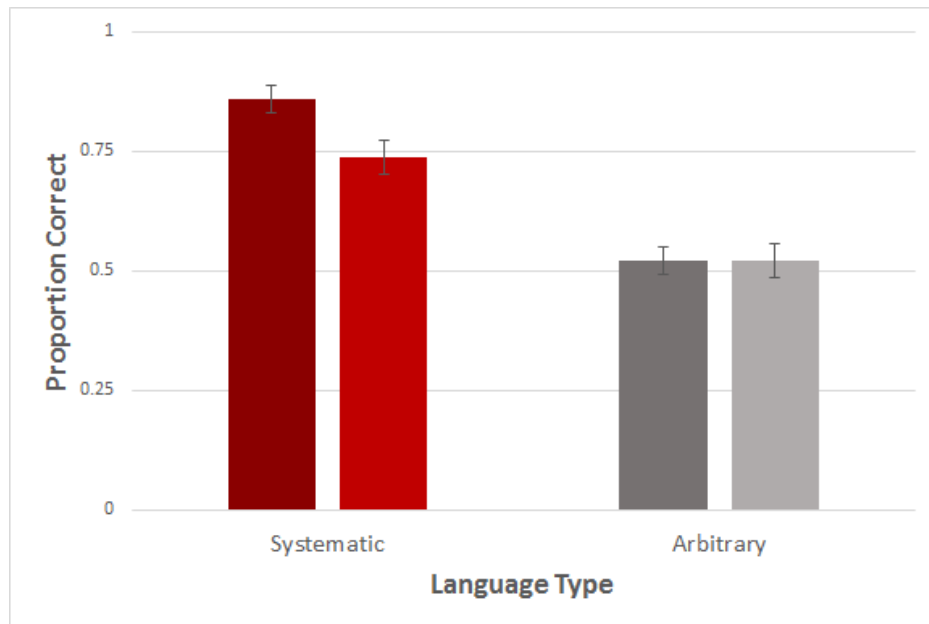


Figure 2.06- A comparison of Categorisation performance between Monaghan et al. (2011)- Experiment 1 (dark), and Experiment 1 (light) presented here. Error bars represent standard error.

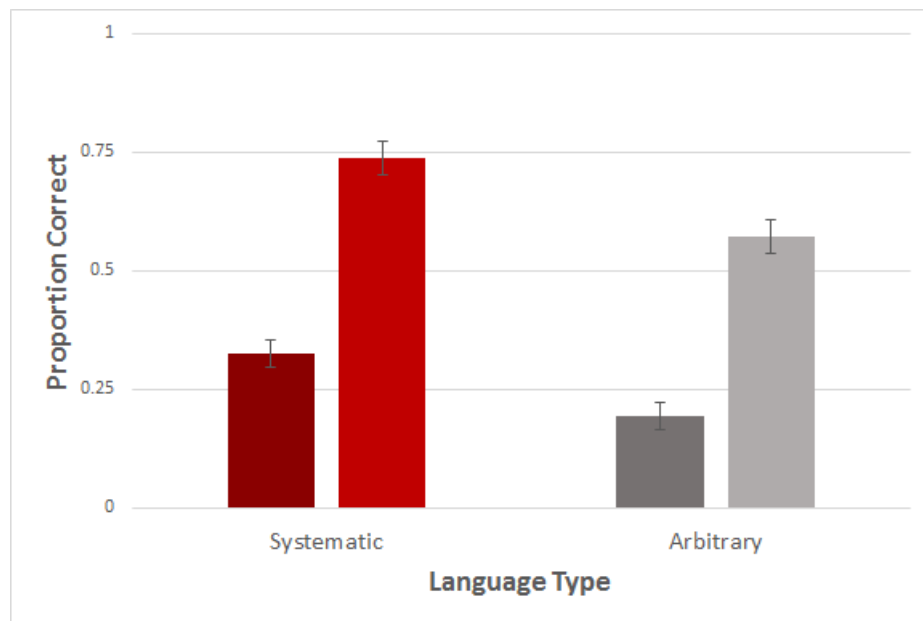


Figure 2.07- A comparison of Individuation performance between Monaghan et al. (2011)- Experiment 1, and Experiment 1 presented here. Error bars represent standard error.

The comparison presented in the above graph in terms of individuation is, however, misleading: as I have acknowledged, my results suggest that systematic language learners actually individuated relatively poorly, and that their individuation results can be better explained by having learned to categorise. This raises a question: how much of the individuation data in Monaghan et al. is due to the same feature? Comparing individuation data for arbitrary language learners between the two conditions is similarly difficult: given a response where participants were only guessing, my participants would guess correctly on 50% of the trials, while Monaghan et al.'s would guess correctly on 8.33% of trials.

For a better comparison, I subtracted the effect attributable to guessing from the observed values for both experiments. For arbitrary languages where no category information is available, this required simply subtracting the average correctness due to chance (50% for Experiment 1 presented here, 8.33% for Monaghan et al.). For systematic languages this meant establishing the baseline correctness from the categorisation score, then subtracting that value from the observed individuation (Experiment 1: 0.79 (observed) - 0.72 (expected) = 0.07 ; Monaghan et al.: 0.325 (observed) - 0.1433 (expected: 84% categorisation/6 options) = 0.182). The comparison of individuation between the two experiments given those values is shown below in Figure 2.08.

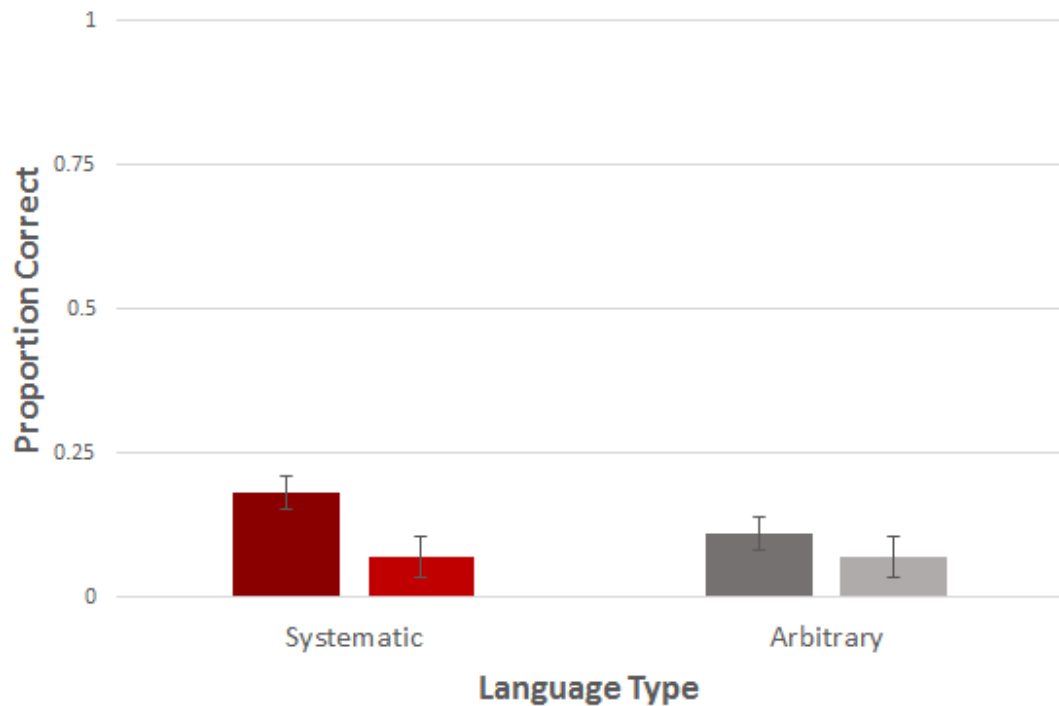


Figure 2.08- A comparison of Individuation performance between Monaghan et al. (2010)-Experiment 1, and Experiment 1 presented here, corrected for the influence that guessing has on performance. Error bars represent standard error.

This correction for the influence of guessing brings the findings of the two experiments relatively close in line with each other; although as mentioned comparing them statistically is inappropriate. Still, both support the same general conclusion: systematicity is good for categorisation, but can be bad for individuation. The signal detection paradigm I used in Experiment 1 allowed me to demonstrate this point even more clearly, and suggested the possibility that

Monaghan et al.’s results should similarly be analysed to account for the effect that learning the category only has on individuation.

The overall positive correlation that I observed between contrastiveness and d' suggests, however, that contrastiveness generally contributes to learning. The pressure for learnability should favor mappings between words and meanings that are more contrastive. In Experiment 1, and in Monaghan et al. (2011), contrastiveness is predicted by language type, but this is likely to be partially due to the fact (especially in Monaghan et al.) that words are chosen from such a small signal space. However, it is possible to maintain a systematic mapping between words and meanings that does not result in words being as similar to one another as in Experiment 1. Below, in a second experiment, I explore this possibility- extending the design of Experiment 1 using a set of labels that is more contrastive.

2.4 Experiment 2

One of the potential problems with extending the findings of Monaghan et al. (2011)’s experiment 1, and my own Experiment 1, presented above, is that it only captures the tension between the learnability benefit for systematicity vs. the learnability penalty due to loss of contrastiveness at a narrow range of values of contrastiveness. That is, in the case of a systematic language for Monaghan et al., the labels ‘fiz’, ‘flz’ and ‘zIf’ for example are very similar to one another, and also relatively similar to the second type of words (‘ga:k’, ‘ka:g’, ‘ku:k’) (how similar

these words are both within each type and between types depends on what model of similarity I use- see Chapter 3). The same is true of the labels used above for Experiment 1, ‘kekete’, ‘kepipe’, and ‘tekipe’ are all quite similar to one another, and although words of the second type are constructed from a different set of phonemes, they are still similarly structured (‘lomumu’, ‘molulo’, ‘mulomo’, etc.).

It is, however, possible to manipulate both the contrastiveness within and between types of words that are used to mark categories systematically. Here, I present the results of a second experiment using a set of words that, rather than making a systematic distinction based on phoneme features, uses two kinds of words that vary in their structure (monosyllables vs. trisyllables) and are additionally maximally distinct within those categories in terms of their phoneme structure. Based on the maximal contrastiveness of the set of experimental stimuli used here for Experiment 2, I predict that the benefit of systematicity will allow for increased learnability for systematic languages without as large of a concomitant reduction to learnability based on contrastiveness.

2.4.1 Methods

Participants

Participants were 28 students (11 female) recruited from the general population of the University of Edinburgh, and were compensated 2.00 GBP for the 15 minutes

required to complete the task. All participants were monolingual English speakers between 17 and 31 years of age. Ethical approval was obtained from the University of Edinburgh in line with BPS guidelines, and informed consent was obtained from all experimental participants.

Participants were assigned randomly to each of 3 experimental conditions.

Conditions 1a (n=6) and 1b (n=7) were counterbalanced systematic language conditions. We found no differences between participants in these subconditions (i.e. it did not matter whether animals were paired with mono or trisyllables), so those subconditions were collapsed for further analysis. The remaining participants (n=15) were assigned to Condition 2 (arbitrary language).

Experimental Design

Label Stimuli

Two lists of twelve nonsense words were created (using the English Lexicon Project Website: Balota et al., 2007): The words followed English phonotactics and were all stressed on the first syllable, but they varied according to the number of syllables (1 vs. 3). The two categories of labels (monosyllables and trisyllables) were selected not only to be distinct from one another, but also to be contrastive within categories (see Table 2.03). Acoustic stimuli for each of the words was created using Apple talk with

the female voice Victoria. As in Experiment 1, I did my best to ensure that the pronunciations were in line with the IPA representation shown below.

Experiment 1		Experiment 2	
Type 1	Type 2	Type 1	Type 2
Systematic	kekete	lomolo	grɒm leɪ
	kekiti	lomumu	dʊr
	kepipe	lumomo	fel ɪt ɪk
	kepipi	molulo	brɒm e di
	kikite	molumo	ref en gəʊr
	kipete	momomo	ʒmen tɪ gɒt
	tekete	momulu	stɪb lɪ kɪl
	tekipe	mulomo	twɪs paɪr ɪm
Arbitrary	kepipi	kikite	skəd
	kekiti	tekipe	blɒn
	kipete	kepipe	mɔɪb
	tekete	kekete	nʌndʒ
	molulo	lomolo	ki:sp
	molumo	momomo	pɪlt
	mulomo	lumomo	skəd
	lomumu	momulu	grɒm leɪ dʊr
			ʒmen tɪ gɒt
			fel ɪt ɪk
			stɪb lɪ kɪl
			brɒm e di
			twɪs paɪr ɪm
			ref en gəʊr
			ʒel er ʊn

Table 2.03- A comparison of the word stimuli used in Experiment 1 vs. Experiment 2. Word stimuli for Experiment 2 are both more different between types (monosyllables vs. trisyllables) and within types.

Image Stimuli

The image stimuli used in Experiment 2 were identical to those used in experiment 1.

Procedure

The procedure used for Experiment 2 was identical to that used in Experiment 1.

Data Analysis

Data analysis for Experiment 2 was identical to that of Experiment 1. d' scores were calculated for each experimental participant and performance between the two conditions was compared using a two sample t-test. Additionally, I performed a repeated measures analysis of variance on the effect of language type (systematic vs. arbitrary) and trial type (target, in-class distractor, out-of-class distractor) on performance. I compared performance between the two conditions with a generalised linear model that included experiment, condition, and trial type as factors. Because the words used in Experiment 2 were not of the same length, they were not amenable to being compared via their Hamming distances, and thus I include no analysis for contrastiveness for Experiment 2.

I predicted that overall the languages used in Experiment 2 would be more easily learned, and that systematic language learners would demonstrate a benefit for systematicity aiding categorisation, but no concomitant learnability penalty on in-class-distractor trials.

2.4.2 Results

Signal Detection

Participants in both the systematic and arbitrary language learning conditions both performed at rates above chance (Systematic d' : $M= 3.37$, $SE= 0.47$;

$t(12)=6.15, p<0.001$; Arbitrary d' : $M=2.36$, $SE= 0.29$; $t(14)= 6.35$, $p<0.001$).

Participants in the systematic condition did not perform significantly better overall than those in the arbitrary condition, although there was a marginal effect ($t(25)= 1.85$, $p=0.068$; Figure 2.09).

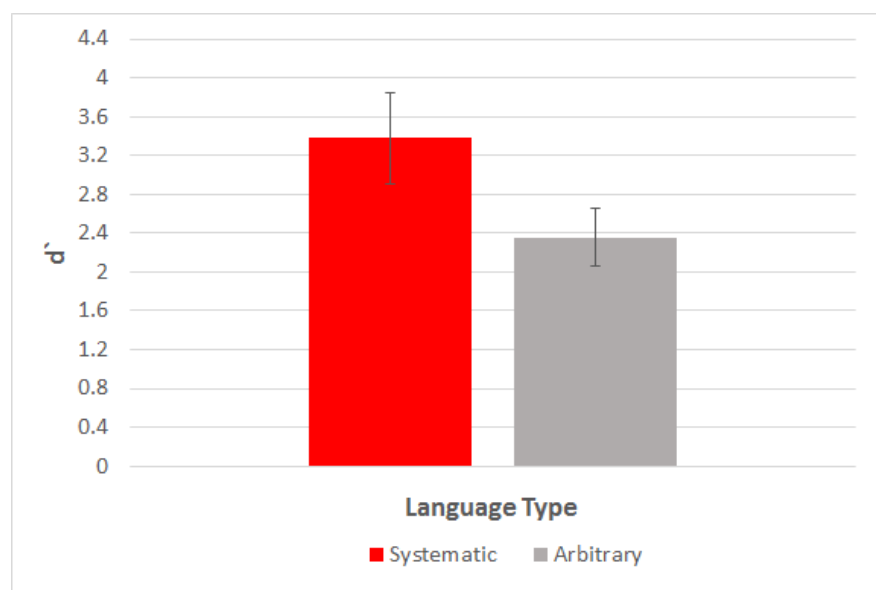


Figure 2.09- d' performance by participants in the systematic and arbitrary conditions of Experiment 2 scored by their ability to identify pairs of objects and labels that they had previous learned in the training phase of the experiment. Performance in both conditions was significantly better than chance ($ps<0.001$). The performance of participants in the two conditions was only marginally significantly different ($p=0.068$).

Repeated Measures Analysis of Variance

The repeated measures analysis of variance showed no significant main effect of condition, in line with the d' omnibus test (Systematic: $M=0.897$, $SE=0.021$; Arbitrary: $M=0.851$, $SE=0.019$; $F(1,83)=2.38$, $p=0.135$). There was however a significant main effect of trial type ($F(2,83)=7.33$, $p=0.0016$): post-hoc comparison using the Tukey-Kramer Multiple comparison test showed that participants performed equally well on target ($M=0.846$, $SE=0.019$) and in-class distractor ($M=0.847$, $SE=0.019$), but significantly better than both on out-of-class distractor trials ($M=0.929$, $SE=0.013$).

There was also a significant interaction between experimental condition and trial type ($F(2,83)=3.93$, $p=0.026$; Figure 2.10). Post hoc analysis of this interaction showed that participants in the arbitrary language learning condition did not perform significantly differently on the three trial types (Target: $M=0.813$, $SE=0.027$; In-Class Distractor: $M=0.863$, $SE=0.027$, Out-of-Class Distractor: $M=0.88$, $SE=0.019$; $F(2,27)=1.87$, $p=0.17$). Participants who learned systematic languages however performed significantly differently depending on trial type ($F(2,23)=9.6$, $p<0.001$): Systematic language learners performed better on out-of-class distractor trials ($M=0.98$, $SE=0.019$) than on either target trials ($M=0.88$,

SE=0.027) or in-class-distractor trials (M=0.83, SE=0.027), which they performed equally well on.

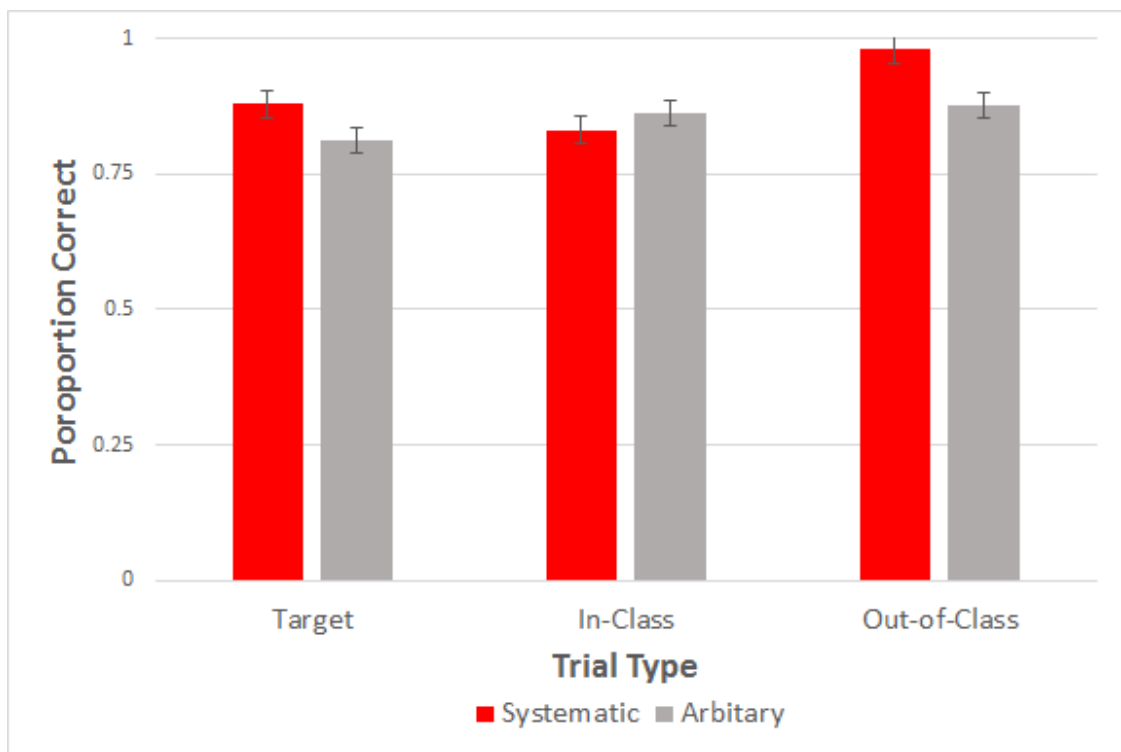


Figure 2.10- Effect of the interaction of experimental condition and trial type on the proportion of correct responses. Participants in the arbitrary language condition performed equally well regardless of the type of experimental trial, while systematic language learners performed best on out-of-class distractor trials.

Additionally, I compared the performance on each trial type between the two conditions using further rmANOVAs. I found that participants performed equally well on both target trials (Systematic M= 0.88, Arbitrary M=0.81; $F(1,420)=2.12$, $p=0.16$) and in-class distractor trials (Systematic M= 0.83, Arbitrary M=0.86;

$F(1,420)=0.43$, $p=0.52$), and better on out-of-class distractor trials than did arbitrary language learners (Systematic $M=0.98$; Arbitrary $M=0.88$; $F(1,27)= 12.62$, $p=0.0015$).

Comparison of Experiments

A repeated measures analysis of variance comparing performance on the two experiments found a significant main effect of experiment: Participants in Experiment 2 ($M=0.87$, $SE=0.0087$) performed significantly better than participants in Experiment 1 ($M=0.62$, $SE=0.0091$; $F(1,161)= 136.75$, $p<0.001$). There was also a significant main effect of condition ($F(1,161)= 11.11$, $p=0.0016$), and a significant effect of trial type ($F(2,161)=43.54$, $p<0.001$).

The two way interaction for Experiment x Trial Type was found to be significant ($F(2,161)= 22.74$, $p<0.004$), driven mostly by the fact that performance on in-class distractor trials was much lower for participants in Experiment 1 (see below). The two way interaction of Condition x Trial Type ($F(2,161)= 38.41$, $p<0.001$) was also significant, mirroring the general finding of both experiments that systematic language learners were significantly better on out-of-class distractor trials. Finally, there was no significant interaction of Experiment x Condition ($F(1,161)= 1.38$, $p=0.246$).

The two way interactions described above are actually easiest to see based on the results of the significant 3-way interaction of Experiment x Condition x Trial Type ($F(2,161)=16.33$, $p<0.001$; Figure 2.11).

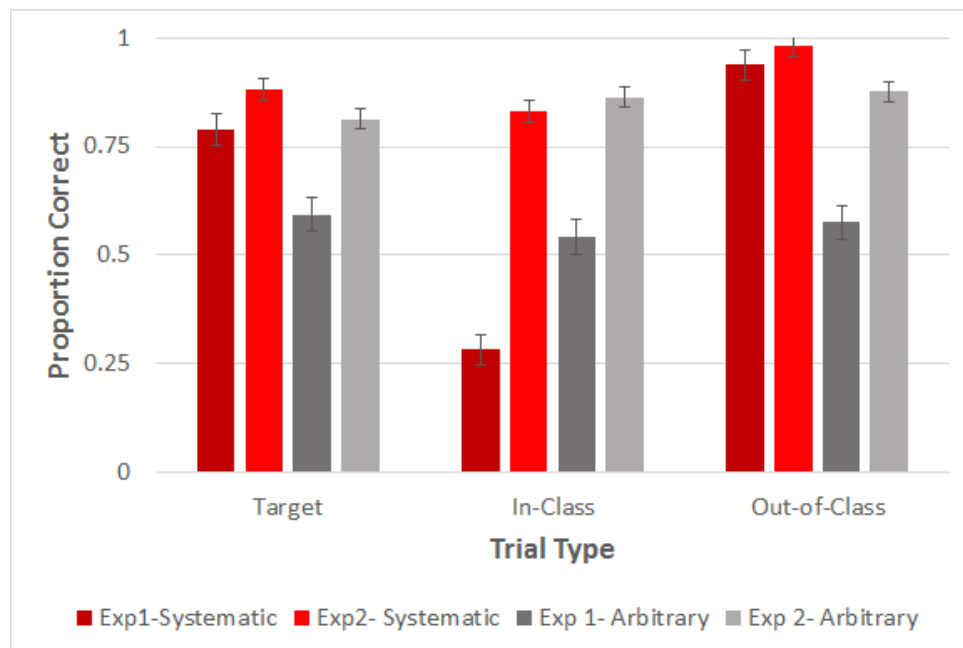


Figure 2.11- Comparison of results between Experiment 1 and Experiment 2.

Post hoc tests showed that systematic language learners performed equally well on both target ($F(1,26)= 3.39$, $p=0.077$) and out-of-class distractor trials ($F(1,26)= 2.03$, $p=0.166$), but that systematic language learners from Experiment 2 performed significantly better than systematic language learners from Experiment 1 on in-class distractor trials ($F(1,26)= 85.85$, $p<0.001$).

2.4.3 Discussion

The results of Experiment 2 dovetail nicely with the suggestion of Gasser (2004) that, all other things being equal, systematic languages are easier to learn when the signal space is more contrastive. Although there was no overall difference in learnability between systematic and arbitrary languages in Experiment 2, systematic language learners performed significantly better on out-of-class distractor trials without the commensurate loss of ability to individuate (lower than chance performance on in-class distractor trials) that we observed in Experiment 1.

The results of Experiment 2 demonstrate that under certain conditions, the costs of systematicity that are incurred by a reduction of contrastiveness can be avoided while still maintaining a benefit for categorisation, and thus suggests that the pressure exerted by the learnability penalty is variable and contingent on contrastiveness. In fact, contrastiveness seems to be a general pressure, as even arbitrary language learners in Experiment 2 performed significantly better than they did in Experiment 1.

Monaghan et al. (2011) also recognized that the pressure for languages to be contrastive would have an important impact on the relative learnability of systematic vs. arbitrary languages. In an additional experiment, they attempted to explore this possibility by creating languages that were still systematic but that came from a larger signal space and were thus more contrastive. Below, I present the

results of a third experiment extending the findings of Monaghan et al. (2011)’s second experiment (actually the third Experiment in their manuscript, but here we will not discuss Experiment 2).

2.5 Experiment 3

In Experiment 2, presented above, we moved from marking categories systematically via phonology to marking categories via word length, Monaghan et al., on the other hand, opted in their 2nd experiment to maintain phonological marking of categories. Rather than creating two types of words that used different phonemes entirely, Monaghan et al. relaxed the constraint that every phoneme in a word should systematically mark the category of that word. So, rather than all 3 phonemes in a cVc trigram being systematic, they created labels where only the coda phoneme in the trigram was systematic (i.e. any word ending in /g/ or /k/ could be used for a noun). Monaghan et al. termed this new marking a ‘half-half’ language, and suggested that systematic marking in that way should provide the benefits of categorisation while decreasing the penalty on individuation (Table 2.04).

Word Type 1				Word Type 2			
fiɜ	fa:ɜ	giɜ	ga:ɜ	ga:k	gik	fik	fa:k
<i>fiɸ</i>	fa:f	gif	ga:f	<i>ga:g</i>	gig	fig	fa:g
flɜ	fu:ɜ	gIɜ	gu:ɜ	gu:k	gIk	flk	fu:k
flɸ	fu:f	gIf	gu:f	gu:g	gIg	flg	fu:g
ɜiɜ	ɜa:ɜ	kiɜ	ka:ɜ	ka:k	kik	ɜik	ɜa:k
ɜiɸ	ɜa:f	kif	ka:f	ka:g	kig	ɜig	ɜa:g
<i>ɜIɜ</i>	ɜu:ɜ	kIɜ	ku:ɜ	<i>ku:k</i>	kIk	ɜIk	ɜu:k
ɜIf	ɜu:f	kIf	ku:f	ku:g	kIg	ɜIg	ɜu:g

Table 2.04- Signal space used for Monaghan et al. (2011) Experiment 2. The first column of each word type shows the signal space from Monaghan et al. experiment 1. Words in that column in bold are those that were actually tested in that experiment. The larger signal space of each word type represents the total available space for Experiment 2, where the systematic difference between the word types can be found only in the coda position of each trigram (but, see below). Finally, cells highlighted in blue show the set of words from the larger signal space that were tested in Monaghan et al.'s experiment 2.

Monaghan et al. (2011) found, in support of their suggestion, that learners of their half-half languages performed best on the task of individuation (better than both arbitrary language learners, and systematic language learners from Experiment 1), and also performed significantly better at categorisation than did arbitrary language learners (Figure 2.12).

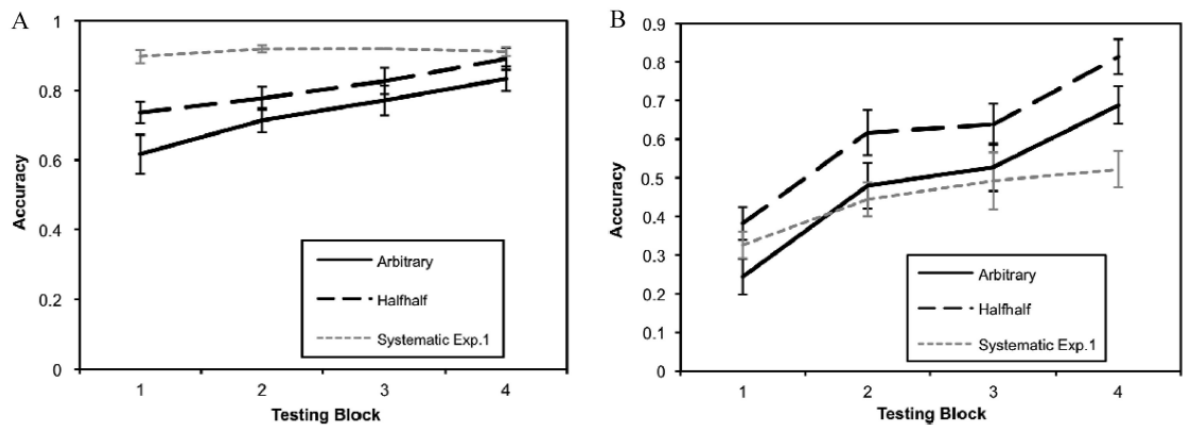


Figure 2.12- *The results of Monaghan et al.'s second experiment, showing a benefit for half-half languages for categorisation (left) and individuation (right).*

By creating half-half languages, Monaghan et al. (2011) explicitly manipulated the size of the available signal space: In experiment 1, with fully systematic cVc trigrams, there were 8 possible words of each type, whereas in experiment 2, with half-half trigrams, there were 32 possible words of each type. In both cases, six words of each type were chosen from the total signal space. However, as we saw in the introduction, size of the signal space and contrastiveness are related, but are not actually the same thing. For example, given Monaghan et al.'s new signal space there would be 906,192 possible ways to select six words from each possible space of 32 words. Monaghan et al., however, test only one of those possible combinations (highlighted in blue in the above table). Within the larger signal space, it is possible to select subsets of words that are relatively more or less contrastive to each other: for example, it's still possible when choosing from the larger signal space to select the exact words used in Experiment 1. To disentangle the effect of contrastiveness

from the effect of signal space size, we extended Monaghan et al.'s half-half language to test random combinations of words taken from the larger signal space, hoping that this manipulation would allow me to trace my results to contrastiveness more broadly, rather than allowing for only a comparison between a single minimally contrastive language (Monaghan et al. Experiment 1) vs. a much more contrastive one (Monaghan et al., Experiment 2). I created Experiment 3 as an attempt to explicitly address this issue.

2.5.1 Methods

Participants

Participants were 60 students (20 male) recruited from the general population of the University of Edinburgh, and were compensated 2.00 GBP for the 15 minutes required to complete the task. All participants were monolingual English speakers between 18 and 33 years of age. Ethical approval was obtained from the University of Edinburgh in line with BPS guidelines, and informed consent was obtained from all experimental participants.

Experimental Design

Where Experiments 1 and 2, presented above, were simple two-way designs, Experiment 3 was created using a 2 (signal space size: large vs. small) x 3 (language type: systematic vs. half-half vs. arbitrary) factorial design.

Label Stimuli

The words used for Experiment 3 were cVcVcV trisyllables, as in Experiment 1 presented above, but the construction of those words was different for participants in each of the six experimental conditions (following the 2x3 factorial design presented above). Large signal spaces were created by combining phonemes similarly to how they were combined in Experiment 1 from a set of six consonants (t, k, p, m, n, l) and four possible vowels (i, e, o, u). At each of the consonant locations, four of the possible six consonants were chosen (2 plosive, 2 sonorant). For large signal spaces, consonants and vowels were combined exhaustively to create 4 possible syllables for each of the first two syllables of the word; for the final syllable only two of the possible four syllables were chosen such that both the consonant and the vowel of that syllable were different. For small signal spaces, all three syllables were created identically to the final syllable for large signal spaces. For systematic languages, plosive consonants were paired with non-rounded vowels, and sonorants consonants with rounded-vowels. For half-half languages, one plosive and one sonorant consonant was chosen for each type of word, and combined with one rounded and one non-rounded vowel (Table 2.05).

All Consonants	Chosen Consonants	Vowels
p	p	i
t	t	e
k	n	o
m	l	u
n		
l		

		Large	Small
Systematic	Type 1	pi pe ti te	pi te
	Type 2	no nu lo lu	no lu
Half-Half	Type 1	pi po li lo	pi lu
	Type 2	te tu no nu	te no

Table 2.05- An example of the creation of syllables for systematic and arbitrary languages that were either large (chosen consonants and vowels combined exhaustively) or small.

This procedure for creating syllables was done for each of the three syllables of the created words, and then those syllables were combined to create total signal spaces from which the words used for each participant could be selected (Table 2.06). Large signal spaces thus resulted in the creation of 32 possible words, while small signal spaces contained 8 possible words.

		Large Signal Space (32 possible words/type)				Small Signal Space (8 possible words/type)			
Systematic	Type 1	Syll1	Syll2	Syll3	Sample Words	Syll1	Syll2	Syll3	Sample Words
		ti	ke	ti	ti ke ti				pi pe te
		te	ki	ke	te ki ke	pi	pe	te	pi ti te
		ke	pi		ke pi ti	ke	ti	ki	ke ti te
	ki	pe		ki pe ke				ke ti ki	
	Type 2	mu	no	lo	mu no lo				mo nu mu
mo		nu	mu	mo nu mu	mo	nu	mu	no lu mu	
no		lu		no lu lo	no	lo	lo	no nu mu	
nu		lo		nu lo mu				no lo lo	
Half-Half	Type 1	pe	no	ke	pe no ke				pi ke pi
		lo	mu	pi	lo mu pi	pi	ke	pi	pi mu ke
		ti	ke		ti ke ke	lo	mu	ke	lo ke pi
		mu	ti		ti mu ti				lo ke ke
	Type 2	pi	mo	lu	pi mo lu				te pi lu
		lu	te	no	lu te no	te	pi	lu	te no no
te		pi		te pi lu	nu	no	no	nu pi lu	
mo		nu		mo nu no				nu no no	

Table 2.06- An example of the creation of words for systematic and half-half languages using large and small signal spaces.

As in Experiments 1 and 2, arbitrary versions of each language were created by randomly splitting the systematic version of that language such that half of the type

1 systematic words were paired with images of one type, and the other half with images of the second type (see Table 2.07).

		Large				Small	
Systematic	Type 1	ti ke ti	te ke ti	ke ke ti	ki ke ti	pi pe te	
		ti ke ke	te ke ke	ke ke ke	ki ke ke		
	Type 2	ti ki ti	te ki ti	ke ki ti	ki ki ti	pi pe ki	
		ti ki ke	te ki ke	ke ki ke	ki ki ke	pi ti te	
		ti pi ti	te pi ti	ke pi ti	ki pi ti	pi ti ki	
		ti pi ke	te pi ke	ke pi ke	ki pi ke	ke pe te	
		ti pe ti	te pe ti	ke pe ti	ki pe ti	pe pe ki	
		ti pe ke	te pe ke	ke pe ke	ki pe ke	ke ti ti	
						ke ti ki	
		mu no lo	mo no lo	no no lo	nu no lo	mo nu mu	
Half-Half	Type 1	mu no mu	mo no mu	no no mu	nu no mu	mo nu lo	
		mu nu lo	mo nu lo	no nu lo	nu nu lo	mo lo mu	
	Type 2	mu nu mu	mo nu mu	no nu mu	nu nu mu	mo lo lo	
		mu lu lo	mo lu lo	no lu lo	nu lu lo	no nu mu	
		mu lu mu	mo lu mu	no lu mu	nu lu mu	no nu lo	
		mu lo lo	mo lo lo	no lo lo	nu lo lo	no lo mu	
		mu lo mu	mo lo mu	no lo mu	nu lo mu	no lo lo	
Half-Half	Type 1	pe no ke	lo no ke	ti no ke	mu no ke	pi ke pi	
		pe no pi	lo no pi	ti no pi	mu no pi	pi ke ke	
	Type 2	pe mu ke	lo mu ke	ti mu ke	mu mu ke	pi mu pi	
		pe mu pi	lo mu pi	ti mu pi	mu mu pi	pi mu ke	
		pe ke ke	lo ke ke	ti ke ke	mu ke ke	lo ke pi	
		pe ke pi	lo ke pi	ti ke pi	mu ke pi	lo ke ke	
		pe ti ke	lo ti ke	ti ti ke	mu ti ke	lo mu pi	
		pe ti pi	lo ti pi	ti ti pi	mu ti pi	lo mu ke	
		pi mo lu	lu mo lu	te mo lu	mo mo lu	te pi lu	
Half-Half	Type 1	pi mo no	lu mo no	te mo no	mo mo no	te pi no	
		pi te lu	lu te lu	te te lu	mo te lu	te no lu	
	Type 2	pi te no	lu te no	te te no	mo te no	te no no	
		pi pi lu	lu pi lu	te pi lu	mo pi lu	nu pi lu	
		pi pi no	lu pi no	te pi no	mo pi no	nu pi no	
		pi nu lu	lu nu lu	te nu lu	mo nu lu	nu no lu	
		pi nu no	lu nu no	te nu no	mo nu no	nu no no	

Table 2.07- An example of a set of possible signal spaces for Experiment 3.

Although words for experiments 1 and 2 were created by using Apple talk to produce whole words (e.g. ‘penoke’) in Experiment 3 we instead used Apple talk to produce only each of the possible cV syllables (e.g. ‘pe’, ‘ki’, ‘ti’, etc.). This change allowed for more precise control over the synthesizer’s pronunciation of the vowel sounds. On each experimental trial, participants were exposed to each word as a set of three syllables following each other immediately. Thus, these words were unstressed and had no confounds from co-articulation. As a cost, however, they sounded more artificial than the words in previous experiments.

Image Stimuli

Images used in Experiment 3 were identical to those used in Experiment 1 and 2.

Contrastiveness

Just as in Experiment 1, a contrastiveness value was calculated for each word learned by every participant, compared to all of their other learned words. This contrastiveness value was included in the analysis of the experiment as a factor in the main repeated measures ANOVA.

Procedure

The procedure for Experiment 3 was identical to Experiments 1 and 2, presented above, but, following Monaghan et al. (2011) participants were only taught 6 words of each type (as opposed to the 8 of each type from Experiments 1 and 2).

Data Analysis

Data analysis for Experiment 3 was conducted similarly to the previous two experiments. The omnibus test of performance using d' as a metric was conducted as a general linear model with my 2 (signal space size) x 3 (language type) factorial design. I predicted a main effect of signal space size, with languages taken from larger signal spaces being easier to learn, in addition to a main effect of language type, with half-half languages being learned the best overall (following Monaghan et al.'s findings).

Additionally, I performed a repeated measures analysis of variance using signal space size (large vs. small) and language type (systematic vs. arbitrary vs. half-half) as between subjects factors and trial type as a within-subjects factor. I predicted that the rmANOVA results would demonstrate the same main effects as the d' analysis, but that there would also be a significant interaction of language type and trial type that would account for the overall superiority of the half-half languages for learning: half-half and systematic language learners would perform

approximately equally well on out-of-class distractor trials, but half-half language learners would perform better on in-class distractor trials due to the greater contrastiveness of their language. Finally, I predicted a further interaction of Size, Systematicity, and Trial type where learners of large systematic languages would perform better on in-class distractor trials than learners of small systematic languages due to the increased contrastiveness of the larger sets of labels.

To explore contrastiveness directly, I performed a linear regression on the d' scores of participants against the average contrastiveness of their language. Additionally, I performed a logistic regression comparing performance on individual trials as a function of the contrastiveness of individual words.

2.5.2 Results

Signal Detection

In line with my predictions, I found a significant main effect of systematicity ($F(2,54) = 6.98, p = 0.002$) but no main effect of signal space size ($F(1,54) = 1.29, p = 0.26$) and no interaction of systematicity \times signal space size ($F(2,54) = 0.36, p = 0.698$). Post-hoc analysis using a Tukey-Kramer test revealed that systematic language learners ($M = 1.53, SE = 0.16$) performed significantly better than arbitrary language learners ($M = 0.66, SE = 0.16$); Half-half language learners ($M = 1.13,$

SE=0.16) performed between systematic and arbitrary language learners, but were not significantly different from either (Figure 2.13).

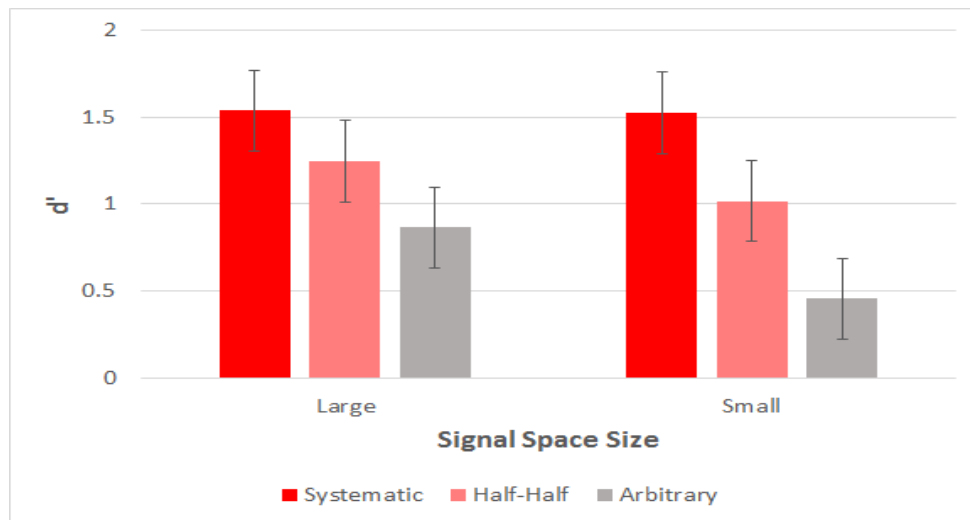


Figure 2.13- The results of the analysis of Experiment 3 using a 2x3 factorial design with d' as the dependent variable. The graph shows a significant main effect of systematicity: systematic language learners performed significantly better than did arbitrary language learners.

Repeated Measures Analysis of Variance

The rmANOVA of correctness did not match up with my hypotheses: I found no significant main effects of lexicon size ($F(1,179) = 0.83$, $p=0.365$) and only a marginally significant main effect of systematicity ($F(2,179)=2.47$, $p=0.094$). There was a significant main effect of trial type ($F(2,179)= 20.87$, $p<0.001$) in line with my predictions.

There was no significant two way interaction of Size x Systematicity ($F(2,179)= 0.33$, $p=0.72$) or of Size * Trial Type ($F(2,179)=0.52$, $p=0.59$). There was however a

significant interaction of Systematicity x Trial Type ($F(4,179)= 9.57, p<0.001$). The three way interaction was not significant ($F(4,179)= 0.83, p=0.51$).

Post hoc analysis of the main effect of trial type using the Tukey-Kramer multiple comparison test showed that participants performed significantly better on target ($M=0.71, SE=0.024$) and out-of-class distractor trials ($M=0.75, SE=0.017$) than they did on in-class-distractor trials ($M=0.56, SE=0.024$).

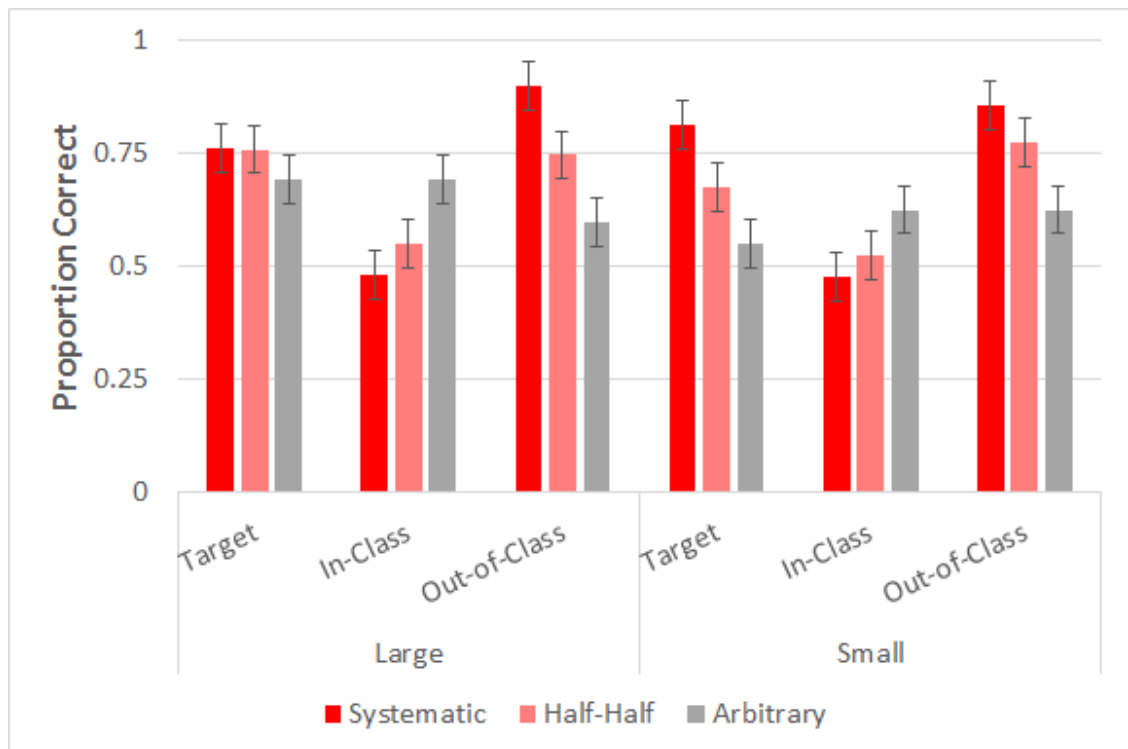


Figure 2.14- The results of Experiment 3, plotting proportion correct and the influence of signal space size, systematicity, and trial type. The graph shows a significant main effect of trial type, and a significant interaction of trial type x systematicity.

Post hoc analysis of the interaction between systematicity and trial type showed that participants in the arbitrary language learning conditions did not perform significantly differently on each of the three trial types ($F(2,59)=0.89$, $p=0.42$). Participants who learned systematic languages, and those who learned half-half languages, however, performed significantly better on Target and Out-of-Class distractor trials than on In-Class distractor trials (Systematic: $F(2,59)=28.92$, $p<0.001$; Half-Half: $F(2,59)=6.77$, $p=0.003$).

Contrastiveness

The results of the linear regression between average contrastiveness and d' revealed no significant correlation between the two ($r=-0.03$, $p=0.99$). No correlations for any language type were significant (all $ps>0.90$).

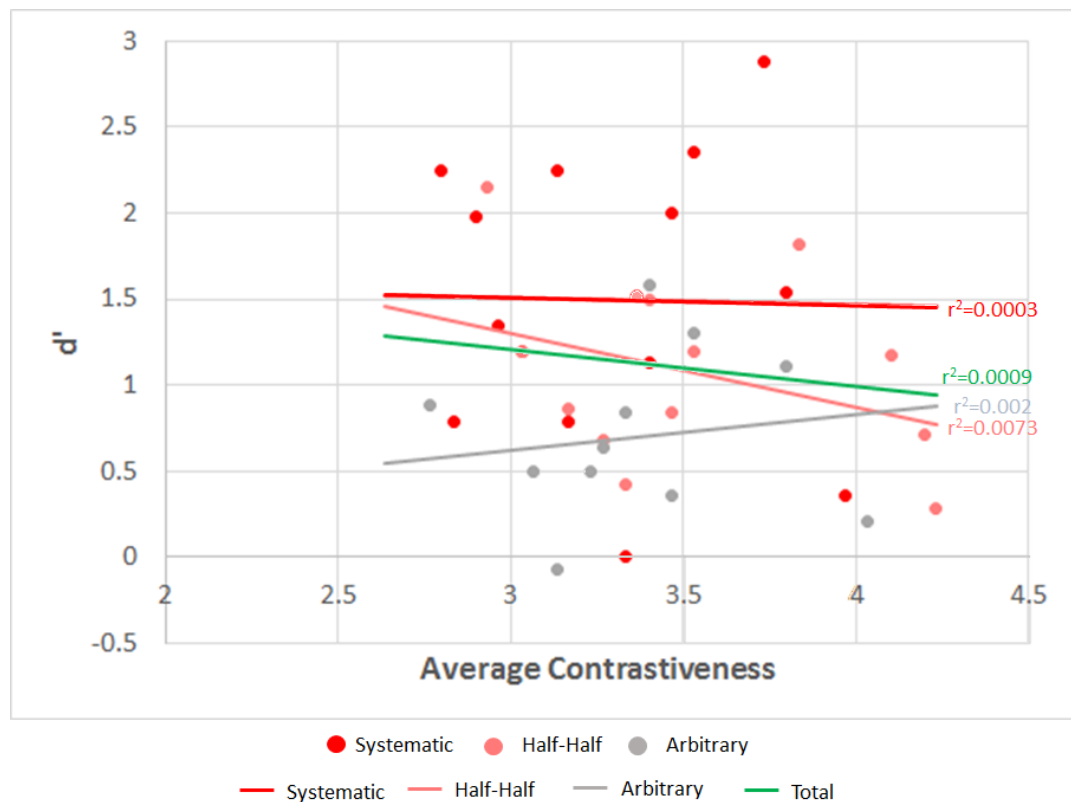


Figure 2.15- The results of a linear regression between d' and average contrastiveness shows no significant correlation between the two ($p=0.995$).

A logistic regression comparing performance on individual trials as a function of contrastiveness was similarly not significant ($r=0.196$, $p=0.28$).

2.5.3 Discussion

The results of Experiment 3 are somewhat confusing: on the one hand, the basic findings of the previous two studies are supported: systematic language learners perform best, with the bulk of this effect driven by their increased competence on out-of-class distractor trials, which they are able to reject correctly with high rates

of accuracy; additionally, as expected, learners of systematic lexica again struggled with in-class distractors, making false alarms at rates around chance (suggesting they were forced to guess on such trials). On the other hand, I did not replicate Monaghan et al.'s findings that half-half languages were easier to learn than fully systematic ones: Although half-half language learners did not perform significantly worse than systematic language learners overall, they also didn't perform better than arbitrary language users. Broken down over trial types, the differences between systematic and half-half language users were again non-significant, suggesting that any benefit for half-half language learners is not manifest in this data.

The manipulations used in this experiment were designed to demonstrate that there is a difference between size of the signal space and contrastiveness. In both of their experiments, Monaghan et al. selected a single language from the possible signal space for each experimental condition to be used for all participants. The words chosen for Monaghan et al.'s experiments were not, however, random: they were actually selected in such a way that they were maximally contrastive for their signal space. To demonstrate this fact, I simulated the process of choosing labels from the entire signal space available for the systematic language used in Monaghan et al.'s first experiment and for the half-half language used in Monaghan et al.'s second experiment. From these signal spaces (8 total words for the Systematic language from Experiment 1; 32 total words for the Half-half language from Experiment 2) I simulated the process of choosing labels over 10,000 runs, with

each run calculating the contrastiveness of those classes of labels both within-class and between-class. The results of my simulation of this process show that the choice of labels used by Monaghan et al. in both experiments was made in such a way that those labels were maximally contrastive. The systematic language used in Experiment 1 had a within-class contrast (the similarity of each set of chosen words to each other) of 1.50, while the average set of words chosen from that signal space had a within-class contrast of 1.426. The different for the half-half language used in Experiment 2 and the average was even larger: the half-half language used was one of the combinations with the highest possible contrastiveness (1.94), compared to the average within-class contrastiveness of 1.72 (Table 2.08).

	<i>Systematic</i>		<i>Half-Half</i>	
	<i>Used</i>	<i>Average</i>	<i>Used</i>	<i>Average</i>
<i>Within</i>	1.5	1.426	1.94	1.72
<i>Between</i>	3	3	2.53	2.5

Table 2.08- *Contrastiveness of systematic label types from Monaghan et al.'s Experiment 1, and half-half label types from Monaghan et al.'s Experiment 3. In each case the set of labels chosen for each type is maximally contrastive for one chosen from that possible signal space.*

The comparison between the learnability of these two languages tells us, as did the comparison between my own Experiments 1 and 2, presented above, that systematic languages that can maintain maximal contrastiveness are easier to learn than systematic languages that cannot. The lack of a benefit for half-half marking in the Experiment 3 data, however, points out that simply increasing the size of the

potential signal space is not enough: the languages constructed from that space also require some optimisation process working over them to ensure that more contrastive languages are selected from the possible combinations. In the case of Monaghan et al. (2011), this was achieved by the deliberate selection of maximally contrastive lexica from the available signal space, but real languages do not have this luxury. Fortunately, the process of iterated learning seems to be exactly the kind that can lead to the emergence of more optimal systems with respect to learnability, transmissibility, and communicative function. Over time, whatever permutation of the lexicon was chosen originally, the process of cultural transmission through iterated learning could slowly move the language towards one of the many possible local optima.

Unfortunately, my own attempts to trace performance directly to contrastiveness in Experiment 3 were unsuccessful. The most likely explanation for this might be that because of the very large number of permutations of chosen lexica, combined with the substantial individual differences displayed by experimental participants, there was really no reason to expect that the data available would be adequate to look at fine-grained distinctions in contrastiveness. In Chapter 5 of this dissertation, I return to an experimental manipulation designed to explore contrastiveness at a more fine-grained level.

One additional feature of the way that Monaghan et al.'s word sets for Experiment 2 were chosen is interesting and potentially problematic: at the level of individual phonemes the half-half language is indeed arbitrary with respect to their first consonant and vowel (e.g. /f/, /g/, /k/, and /ʒ/ all appear as the first consonant; and /I/, /i/, /a:/, and /u:/ all appear as the first vowel), but considered more distantly their chosen half-half language could actually be considered systematic. Type 1 words from Monaghan et al.'s Experiment 2 begin with 'fi-', 'ʒi-', 'ʒ -', 'gu:-', 'ka:-', and 'ku:-' and always end with /f/ or /ʒ/. Type 2 words on the other hand begin with 'ga:-', 'gu:-', 'ka:-', 'fi-', 'fI-', and 'ʒi-' and always end with /g/ or /k/. Because neither the word beginnings (when considered as segments (onset-rhyme pairs), rather than individual phonemes) nor the word endings overlap between these two languages, they can be considered, in some sense, to be fully systematic, rather than only half-half.

Unfortunately, some of the lexica constructed for my own half-half languages fell prey to the same peculiarity, take for example the following half-half language:

Type 1	Type 2
pe ki pi	ti te lu
pe ki ke	ti te no
pe nu pi	ti mo lu
pe nu ke	ti mo no
lo ki pi	mu te lu
lo ki ke	mu te no
lo nu pi	mu mo lu
lo nu ke	mu mo no

Table 2.09- A sample language from Experiment 3

By my rules for constructing a small half-half language these labels make sense - possible syllables were constructed randomly from the set of: 'pe', 'pi', 'ke', 'lo', 'lu', 'mo', and 'mu' and then these syllables were chosen for each syllable locus such that at each locus for each type of label two of the syllables would be used. However, this ends up with a strangely mixed almost fully systematic language. In terms of consonants, the language ends up being almost fully systematic: type 1 labels use the consonants /p/, /l/, /k/, and /n/, whereas type 2 labels use the consonants /t/, /m/, /n/, and /l/: between the two types /l/ is the only consonant seen in both word types. Looking at the Hamming distance of these label types compared to one another, the difference becomes even more glaring, with the out-of-class contrast being 6 across all possible languages selected from this set: at each individual locus of consonant or vowel the same phoneme does not appear in the two languages: i.e., although /l/

appears in both word types, it appears only in the first syllable for type 1 word and only in the final (intentionally systematic) syllable for type 2 words.

Whether these languages should be considered systematic or not is an interesting question, although it seems unlikely that experimental participants would recognize the systematicity of the initial segments in Monaghan et al.'s half-half languages given the much more transparently systematic coda phoneme. Still, whether these kinds of associations can be recognized as systematic and thus aid categorisation learning is an open empirical question: real world corpus studies showing that systematicity at the level of the lexicon is often distribution and statistical in nature (e.g. Monaghan et al., 2014) suggest that this possibility would be interesting to explore further.

The possibility of choosing languages from a large signal space that are incidentally systematic (but not due to being motivated) becomes even greater for larger signal spaces: from the 32 possible words of each type in my large signal spaces there would be many combinations of choosing 6 words of each type randomly that would result in languages that were fully systematic. Some of those languages would be fully systematic in the same way as the systematic languages in the above experiments, while a larger subset would be fully systematic in the way described above.

2.6 General Discussion and Conclusions

In this chapter I have laid out the results of three experiments designed to more fully explore the differential effects of non-motivated systematicity and arbitrariness (non-motivated, non-systematic) on the learnability of artificial languages. The experimental methodologies used here, while utilising a different experimental protocol (signal detection vs. AFC) were modelled after those used by Monaghan et al. (2011), and produced results that were broadly consistent with previous findings: non-motivated systematic mappings between words and meanings facilitate learning, although this effect is largely mediated by the fact that they allow for efficient learning of categories in such a way that out-of class distractors are quickly and easily rejected. Also in line with the findings of Monaghan et al., systematic mappings between words and meanings can produce penalties for learnability when it comes to individuation, and these are easiest to spot in the data by considering the fact that systematic language learners perform at or below chance on in-class distractor trials in each of the three experiments- suggesting that they are learning category structure, rather than learning individual meanings adequately.

Overall then, my results are supportive of the claim that systematicity, when applied to a signal space of a given dimensionality, reduces the average contrastiveness of labels to one another, and thus impairs individuation. Creating a language that is still systematic, but less constraining, results in languages that gain

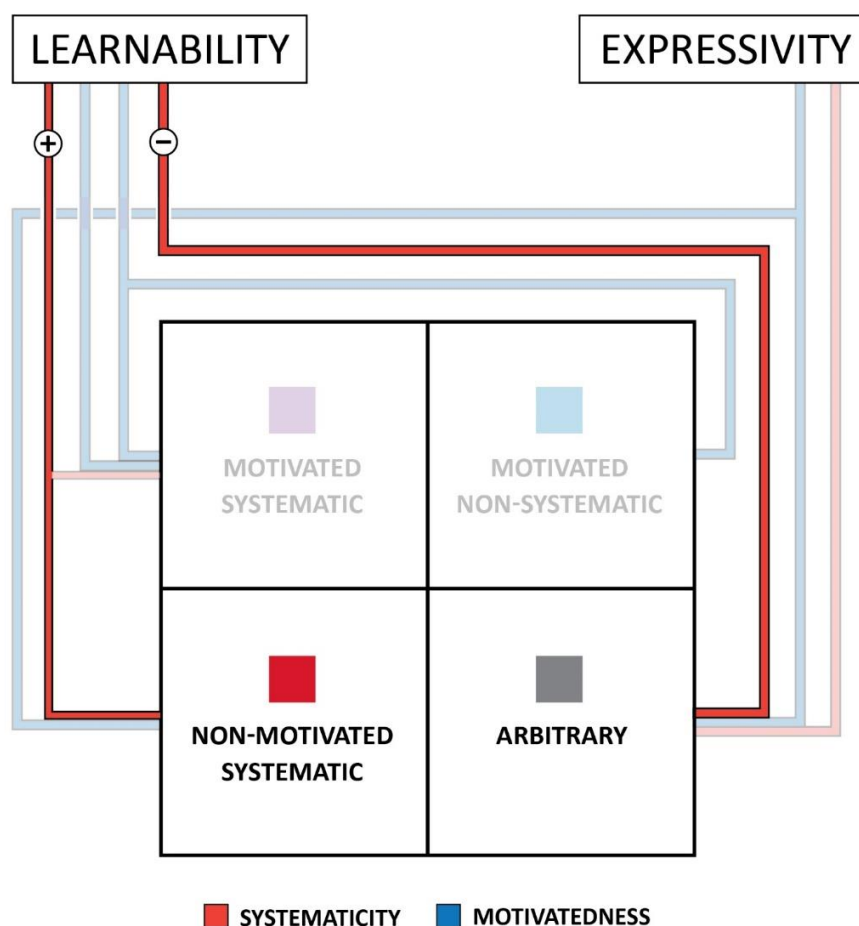
the benefits of both systematicity (i.e. increased categorisation) and arbitrariness (increased individuation). Of note, despite not finding support for the superiority of this half-half language over fully systematic ones in any of the relevant metrics, my manipulation in Experiment 3 revealed that simply modifying the size of the available signal space does not necessarily create conditions in which a lexicon chosen from that space becomes more learnable- it merely creates the possibility for more contrastive lexica to be chosen.

My attempt to manipulative contrastiveness of learned language by selecting randomly from the available signal space was successful, but not predictive of learning. Part of the reason for this may be due to the contrastiveness measure used for the above experiments (Hamming Distance). Hamming distance applies to strings of phonemes, being agnostic to anything about either phonology or psychological reality, suggests that n and p , for example, are as similar to one another as are p and b , which seems unlikely to be the case. For this reason, Monaghan et al. (2011) used a phonological feature encoding for their neural network that captures the relative similarity of certain features to one another. Thus, a phonological feature encoding highlights that systematic configurations of the language rely on mapping similar encodings to similar meanings- not just mapping unrelated sounds to meanings in a systematic way. The degree to which these phonological features are relevant to human perception however is an open question- one that I attempt to address in Chapter 3. Although this question may

seem separate from the central questions of this dissertation, it is actually potentially very important: if we believe that the learnability pressure towards arbitrariness is contingent on the degree of contrastiveness between words, it is crucial to understand what measure of contrastiveness is most relevant to human language learners.

Chapter 3

Systematicity and Learning II: The role of phonological similarity in artificial language learning



***Figure 3.01-** In Chapter 3, I explore the same comparison between languages that are systematic (but non-motivated) and languages that are arbitrary. Here, I focus on the pressure for arbitrariness due to confusability and a discussion of what contrastiveness measure is most relevant for human language learners.*

In Chapter 2, I presented a series of experiments that explored two learnability pressures. The first of these (Learnability pressure A, above) is a pressure towards systematic mappings between words and meanings that are suggested to make learning easier. The results of the experiments presented in Chapter 2 suggests that the increased learnability of systematic languages is because they allow for increased categorisation (recognizing that the word for a given meaning is of the correct category, if not the exact correct word). The second learnability pressure (Learnability Pressure B, above) is one that favors arbitrary associations between words and meanings because systematic associations tend to reduce the size of the available signal space, resulting in words that are, all other things being equal, more similar to each other. In this chapter, I focus on an exploration of this issue of this issue of similarity, which I have characterized as the contrastiveness of a word to other words with similar meanings and to the lexicon more generally. The central question of this chapter is what measure of contrastiveness is most relevant for human language learners: only by understanding the types of similarity that result in increased confusability and a reduction of learning can we understand the strength of the pressure towards arbitrary word-meaning associations.

In addition to a comparison of the ability of different contrastiveness metrics to predict the relative learnability of systematic vs. arbitrary word-meaning associations, I also critically explore the notion of *categorisation* presented in Chapter 2, suggesting that exploring categorisation in the way that we and previous authors have is not actually entirely appropriate.

3.1 Background and Rationale

To rehearse, systematicity refers to isomorphisms between a set of words and a set of meanings such that similar words are mapped to similar meanings. In English and many other languages this is exemplified by phonaestheme clusters: for example in English the ‘gl-’ cluster is found in a number of words associated with light and vision (e.g., ‘*glint*’, ‘*gleam*’, ‘*glare*’). Systematic associations between words and meanings are suggested to be potentially beneficial for language learning since they include a regularity in word-meaning mappings. However, the ability to identify and exploit systematic associations is contingent on previous exposure to other examples of the association. For instance, given the ‘gl-’ phonaestheme example a naïve speaker of English would be unlikely to pair ‘*glare*’ with its specific meaning, but once familiar with a few tokens from the cluster would be more likely to guess the appropriate meaning for a low frequency word like ‘*gloom*’.

The ability of learners to leverage systematic associations between words and meanings for language acquisition is contingent largely on two factors, which we saw

in Chapter 2. Systematic word-meaning mappings take advantage of a similarity between words of a given type, and this means that experience with those words can be generalized (i.e. coming across a new ‘*gl-*’ token makes the meaning easier to guess). However, because words in systematic languages are similar to one another, and stand for similar meanings, they might be more easily confusable, either in terms of learnability or as a function of communicative pressure and transmission error.

In 2011, Monaghan et al. reported a series of experiments and computational models of language learning that they designed to test the effect of systematicity on learning. Their primary finding was that systematic and arbitrary (i.e. neither iconic nor systematic) lexicons facilitate different types of learning. The task of individuating the meaning of a given word (i.e., selecting the appropriate pairing of label and referent), was promoted by arbitrariness, but the process of categorization (i.e., choosing a referent of the appropriate *type* for a given label, but not necessarily the exact referent) was aided by systematic mappings between forms and meanings.

In Chapter 2 of this dissertation, we presented the results of three experiments that extended Monaghan et al. (2011)’s exploration using different stimuli and a new experimental protocol (signal detection rather than alternative-forced-choice). The results of those experiments were broadly supportive of Monaghan et al.’s original findings, although they highlighted that predicting the

costs and benefits of systematicity of language learning can be difficult. The systematic structure used by Monaghan et al. as based on a mapping of phonology to meaning: words for a given type of meaning were similar to one another based on their phoneme use. Words of the first type, for example, used the plosive consonants /g/ and /k/, which differ only in terms of their voicing. Words of the second type, on the other hand, used the fricative consonants /f/ and /z/, which are also similar to one another. The phonemes used in the two word types are, however, much more different between types than they are within: /f/ and /g/ for example differ in terms of their sonority, voicing, degree of stricture, palatalization, roundness, and tongue features (based on a set of 11 phonological features from Harm & Seidenberg, 1999).

The same is true of the phonemes that I used to construct my languages in Experiment 2 of Chapter 2. There, the plosive consonants /p/, /t/, and /k/ are more similar to each other based on their phonological features than they are to the sonorant consonants /m/, /n/, and /l/. Despite the fact that I created two types of words based on phonological features, I found that the simple edit distance between words (Hamming distance) was actually predictive of the learnability of those words. This is somewhat curious, as Hamming distance applied to strings of phonemes is agnostic to their phonological features: /k/ and /g/ differ only slightly in terms on their phonological feature mapping (voicing) compared to the difference between /f/ and /g/ (6 features), but Hamming distance considered the difference between /k/, /g/, and /f/ to be equal.

This raises an interesting question: what kinds of systematic mappings between features of words and features of meanings can language learners take advantage of, and how can we best define a metric of contrastiveness that captures human biases. The comparison of results between the models used in Monaghan et al.(2011), which represent words as clusters of phonological feature dimensions, and the results of their experimental participants (to whom words were presented as auditory stimuli) suggests that phonological similarity is important for establishing systematic associations and also establishing the confusability of words to one another. Computational models that encode words based on their phonological features learn arbitrary and systematic languages similarly to humans, so the suggestion might be that those phonological features similarly capture human learning. However, my results from Chapter 2 suggest that Hamming distance, which is agnostic to phonological features, also predicts human performance – so, which of these is a better measure of the metric along which similarity is established cognitive for human language learners?

3.2 Chapter Outline

In this chapter, we replicate and extend the model used by Monaghan et al. (2011), including a manipulation of phonological clustering. As mentioned, Monaghan et al., used phonemes for each class of label that were similar based on their phonological features- a configuration I refer to as *phonologically clustered*, which is typical for

these kinds of experiments. However, it is possible to create equally systematic languages where this feature of phonological clustering is broken- a configuration I refer to as *phonologically dispersed*. In phonologically dispersed languages, I map dissimilar phonemes to meanings that are similar: for example, /f/ and /g/, despite being quite different, can both be used to construct words for similar meanings. The inclusion of this phonological clustering manipulation allows us to explore the benefits of systematicity more generally, and to test the degree to which a phonological feature encoding is appropriate for the modeling of human perception in a learning task like this one.

Additionally, I explore the effect of phonological clustering on learning abilities of the computational model further by including a second set of phonemes (contrasting voiceless plosives vs. voiced sonorants) which are equally similar within group but more dissimilar between groups. These two manipulations extend Monaghan et al.'s original 2 level (systematic vs. arbitrary) design to a 2 (systematic vs. arbitrary) \times 2 (phonologically clustered vs dispersed) \times 2 (phoneme set) factorial design that allows us to more fully explore the features of systematic associations between words and meanings that give them their learning benefits (for categorisation) and/or accrue them penalties (due to confusability). I find that the model predicts significant main effects of all 3 factors: systematic languages favor categorisation, while the phonological clustering of those languages predicts the loss of individuation for systematic mappings (phonologically dispersed languages are

individuated very well). Finally, the use of plosive vs. sonorant phonemes, because it results in words that are more phonologically distinct between categories, increases the ability of the model to categorise correctly.

In exploring the effects of these experimental manipulations on the learnability of the model, I also discuss the appropriateness of the term *categorisation*, suggesting that: a) neither Monaghan et al. (2011), nor I, actually measure the ability of either the model or human participants to recognize category structures, and b) that, for this reason, what we have called categorisation is not strictly separate from individuation. In discussing this fact, I return to and re-evaluate some of the findings of Monaghan et al.'s experiments, suggesting that some of the claims made about the benefits of systematicity for categorisation may need to be re-evaluated, or at least explored more directly experimentally.

In addition to an extension of the model used in Monaghan et al. (2011), I present the results of an experiment exploring the effect of phonological dispersion on the learning of human participants. Following the experiments presented in Chapter 2, I use a signal detection paradigm for this exploration. Using this manipulation, I find that for human participants the degree of phonological clustering does not have a significant effect on learnability, suggesting that the model's phonological feature representation overestimates the confusability of the phonemes for human participants. This finding allows for a further discussion about

the notion of contrastiveness, and how to best predict the chance that human language learners will be confused by similar words.

3.3 Simulation 1

The simulation that I present here moves from the 2 level (systematic vs. arbitrary) design used by Monaghan et al. (2011) to a more robust 2 (systematic vs. arbitrary) x 2 (phonologically clustered vs. dispersed) x 2 (phoneme set) experimental design.

This allows us to explore the possibility that systematic mappings not based on phonological feature similarity might be learned differently by a model that encodes labels as strings of phonological features, and also to explore the degree to which similarity within groups and dissimilarity between groups influences the learnability of systematic lexica.

We replicate the connectionist model described by Monaghan et al., testing how well input patterns (phonological feature representations) map onto output patterns (which represent meanings).

3.3.1 Methods

Networks

We (myself and Dieuwke Hupkes, a visiting MSc student) replicated the network architecture used by Monaghan et al. (2011). A feed-forward connectionist model

was constructed with either 33 or 66 input units, 20 hidden units, and 10 output units. The input layer was fully connected to the hidden layer, and the hidden layer fully connected to the output layer. Words, which served as input to the network, were either 3-phoneme CvC combinations constructed from Monaghan et al.'s original phoneme set (33 input units), or the 6-phoneme cVcVcV combinations used in Experiment 1 of Chapter 2 (66 input units). Each phoneme was represented by a binary pattern over 11 input nodes (taken from Harm and Seidenberg, 1999: see Table 3.01).

		Consonants													
		Symbol	Sonorant	Consonant	Voice	Nasal	Degree	Labial	Palatal	Pharyngeal	Round	Tongue	Radical	Example	
Vowels	Monaghan	Type 1	f	-0.5	1	-1	-1	0	-1	1	-1	1	0	0	fit
			ʒ	-0.5	1	0	-1	0	-1	0	-1	-1	0	0	beige
		Type 2	g	-1	1	0	-1	1	-1	-1	-1	-1	-1	0	give
			k	-1	1	-1	-1	1	-1	-1	-1	-1	-1	0	kite
	Experiment 2	Type 1	t	-1	1	-1	-1	1	-1	1	-1	1	0	top	
			k	-1	1	-1	-1	1	-1	-1	-1	-1	-1	0	kite
		Type 2	m	0	0	1	1	1	1	0	-1	1	0	0	mop
			n	0	0	1	1	1	-1	1	-1	-1	1	0	not
	Monaghan	Type 1	i	1	-1	1	0	0	-1	0	-1	-1	0	1	heed
			ɪ	1	-1	1	0	0	-1	0	-1	-1	0	-1	hid
		Type 2	u	1	-1	1	0	0	-1	-1	-1	1	0	1	boot
			a	1	-1	1	0	-1	-1	-1	1	-1	-1	-1	hot
	Experiment 2	Type 1	i	1	-1	1	0	0	-1	0	-1	-1	0	1	heed
			e	1	-1	1	0	-1	-1	0	-1	-1	-1	1	paid
		Type 2	o	1	-1	1	0	-1	-1	-1	-1	1	-1	1	toad
			u	1	-1	1	0	0	-1	-1	-1	1	0	1	boot

Table 3.01- The phonological feature codings used in Simulation 1, taken from Harm & Seidenberg (1999).

We generated output patterns (meanings) according to the procedure described in Monaghan et al. (2011). Two category prototypes were generated: for the first category (glossed as “objects”), values for each output unit were initialized at 0.25 and for the second category (glossed as “actions”) they were initialized to 0.75. Six individual output patterns were generated from each prototype activation pattern by randomly changing the values of each of the output units in the range of ± 0.25 : thus, all output units were in the range of 0-0.5 for objects and 0.5 – 1.0 for actions. All meanings from the same category were therefore represented by similar output representations and were distinct from output representations for the other category. A new set of output layer activation patterns were generated at random for each simulation run to avoid the results being biased by a particular set of output layer initialisations that may have either favored or penalized learning.

Experimental Design

Simulation 1 uses a 2 (systematic vs. arbitrary) \times 2 (phonologically clustered vs. dispersed) \times 2 (phoneme set) factorial design.

Label Stimuli

We trained the network to map between input and output representations (forms and meanings respectively) for 12 input-output pairings, corresponding to 12 form-

meaning associations. We generated eight sets of labels, which varied according to a 2x2x2 design (see Table 3.02).

For *phonologically-clustered* languages there were four sets of labels with distinct phonological features. For the conditions using Monaghan et al.'s phonemes, one subset of six labels used fricatives /ʒ/ and /f/ and vowels /i/ and /I/; the second subset of 6 labels used the plosives /g/ and /k/ and the vowels /a/ and /u:/. For the *phonologically-dispersed* languages the two subsets were constructed using the consonants /f/ and /g/ with the vowels /i/ and /ɒ/ (set 1) or the consonants /ʒ/ and /k/ with the vowels /I/ and /u:/. For the conditions using Nielsen & Rendall (2012)'s phonemes, one subset of six labels used plosives /t/ and /k/ and the unrounded vowels /i/ and /e/; the second subset of labels used the sonorants /m/ and /n/ and the rounded vowels /o/ and /u/. For the phonologically-dispersed versions the subsets were constructed using the consonants /m/ and /t/ with the vowels /u/ and /e/ (set 1) or the consonants /n/ and /k/ with the vowels /o/ and /i/. The set of phonemes used to create words of each type for each of these factors is shown in Table 3.02, along with the phonological feature differences for those sets of phonemes.

		Clustered		Dispersed	
		Type 1	Type 2	Type 1	Type 2
Monaghan	Within	f	g	f	3
		3	k	g	k
	Between	3	1	6	5
		5		3.25	
Nielsen	Within	i	u	i	I
		I	a	a	u
	Between	1	5	5	3
		3.5		3	
Nielsen	Within	t	m	m	n
		k	n	t	k
	Between	2	4	8	6
		6.5		4.5	
Nielsen	Within	i	o	u	o
		e	u	e	i
	Between	2	2	4	4
		3		2	

Table 3.02- Phoneme sets used to create words for Simulation 1. The distance both within each set of phonemes and between the two sets of phonemes is given for each set of phonemes.

A comparison of the average Euclidean edit distance for each of the four sets of phonemes can be seen below, in Figure 3.02.

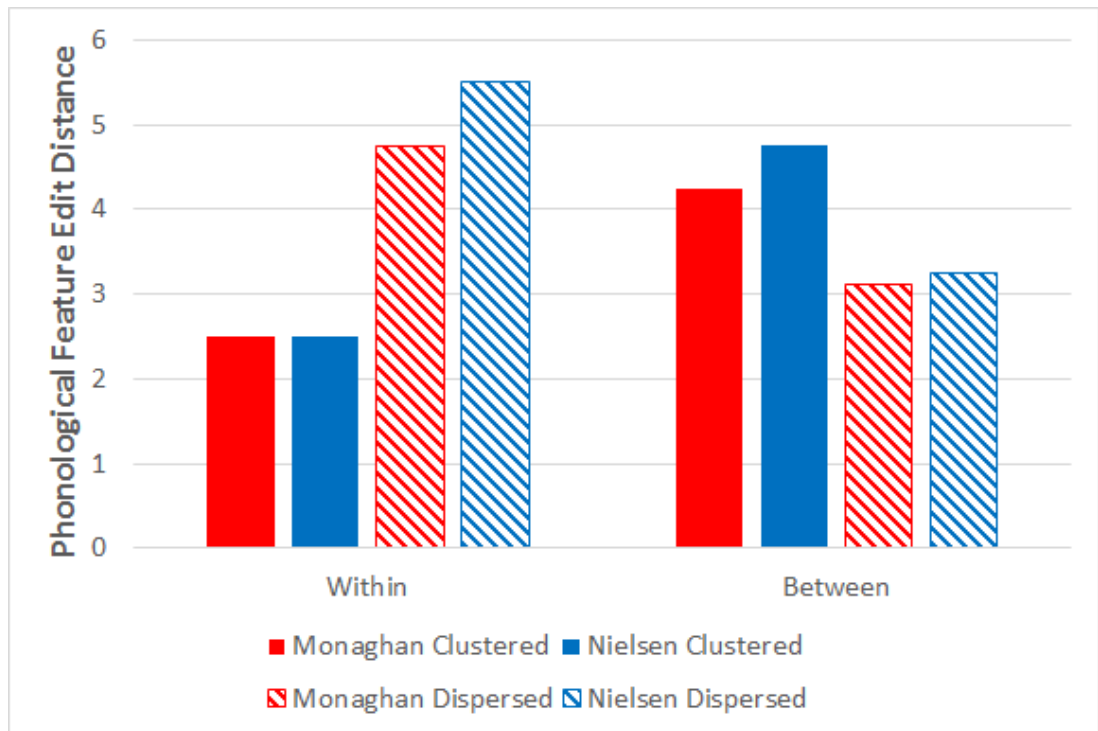


Figure 3.02- The average phonological feature edit distance both within and between the groups of phonemes used to create language for Simulation 1.

The set of phonemes used by Monaghan et al. is slightly more similar than the set used by Nielsen, both in terms of within class (Monaghan= 3.63, Nielsen=4) and between class similarity (Monaghan=3.69; Nielsen=4). This suggests that we should find a significant effect of phoneme inventory, with the languages made from phonemes from Experiment 2 being learned better by the model in terms of both individuation and categorisation than languages made from Monaghan et al.'s original phonemes.

The manipulation of phonological dispersion results in phonemes for phonologically clustered languages that are similar within type ($M=2.5$), but different between type ($M=4.5$). Phonologically dispersed languages, on the other hand, are similar within type ($M=5.13$) but different between type ($M=3.19$).

The manipulations of phoneme inventory and phonological dispersion were crossed with whether the languages were *arbitrary* (3 meanings from each category of meaning were associated with each subset of labels) or *systematic* (all meanings from one category were associated with a single subset of labels). This yields eight language types. In the *systematic phonologically-clustered* lexica, similar sounding words map to similar meanings, and the words within a category have high featural similarity (i.e. they are composed of e.g. fricatives and front vowels or plosives and back vowels). In *systematic phonologically-dispersed* lexica, similar-sounding words map to similar meanings (e.g. all words featuring a /ʒ/ or /k/ will have similar meanings), but words within a category have low featural similarity (e.g. /ʒ/ or /k/ share few phonological features). All four possible arbitrary lexica, whether clustered or dispersed, break this systematicity: similar-sounding words are no more likely to have a similar meaning than not (see Table 3.03).

Euclidean Distances

	Systematic				Arbitrary			
	Clustered		Dispersed		Clustered		Dispersed	
	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2
Monaghan	fiɜ	gak	fig	ɜlk	fiɜ	fIɜ	fig	ɜlk
	ɜif	kag	fag	ɜuk	kug	gak	fag	ɜuk
	fIɜ	guk	gif	kIɜ	ɜIf	ɜif	gif	kIɜ
	ɜIf	kug	gaf	kuɜ	kak	guk	kuɜ	gaf
	fIf	gug	gig	kIk	fIf	ɜiɜ	kIk	gig
Nielsen	ɜIɜ	kak	faf	ɜuɜ	gug	kag	ɜuɜ	fug
	EDW:	2.91	EDW:	3.53	EDW:	4.09	EDW:	3.99
	EDB:	4.72	EDB:	4.17	EDB:	3.78	EDB:	3.79
	mo mo mo	ti ti ti	mu mu mu	no no no	mo mo mo	ti ti ti	mo mo mo	nu nu nu
	mo mo nu	ti ti ke	mu mu te	no no ki	nu nu mo	ke ke ke	mo te mo	ki nu ki
Nielsen	nu nu nu	ke ke ke	mu te mu	ki no no	mo nu mo	ke ti ti	te mo te	nu ki ki
	nu nu mo	ke ke ti	te te te	ki no ki	ke ke ti	nu nu nu	nu nu ki	mo mo te
	nu mo nu	ti ke ke	te te mu	no ki ki	ti ke ke	nu mo nu	ki nu nu	te te te
	mo nu mo	ke ti ti	te mu te	no ki no	ti ti ke	mo mo nu	nu ki nu	te te mo
	EDW:	4.35	EDW:	6.69	EDW:	6.8	EDW:	6.81
Nielsen	EDB:	8.3	EDB:	6.78	EDB:	6.26	EDB:	6.68

Table 3.03- Labels used in Simulation 1.

EDW= Average Euclidean distance within label type.

EWB= Euclidean distance between label types.

As in Monaghan et al. (2011), Euclidean distances between the phonological feature representations for each set of words were calculated (the labels ‘gɒk’ and ‘gɒg’, for example, would share 32 of 33 features and thus their Euclidean distance from each other would be 1) along with the mean Euclidean distances between sets of words in a given condition. This clustering vs. dispersion manipulation shifts the mean Euclidean between the labels associated with a particular category of meaning for

systematic languages. Euclidean distances within and between all categories can be seen in Table 3.03. Additionally, Figure 3.03, below, shows the Euclidean distances visually.

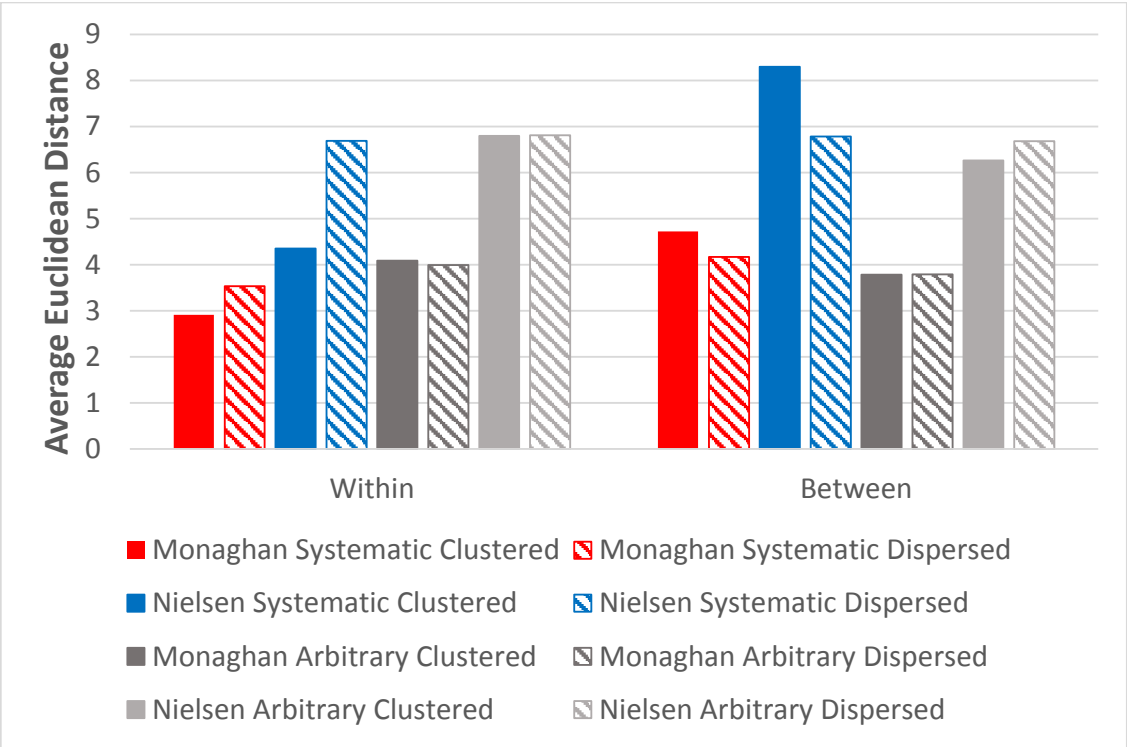


Figure 3.03- *The average Euclidean distance between phonological feature representations of words use in Simulation 1.*

Procedure

Training

Separate networks were trained on all eight language types, manipulating systematicity, phonological dispersion, and phoneme set in a 2x2x2 factorial design.

Weights on connections between units were initially randomized with a uniform distribution in the range of ± 0.5 . The model was trained by back-propagation of error with gradient descend (with a learning rate of 0.05), where after each form-meaning pair was presented the connection weights were adjusted to bring the network's actual output closer to the target output meaning for that pattern. A training block involved the presentation of all 12 input-output pairings in random order and the performance of the model was assessed after 10, 20, 30, and 40 blocks of training.

Testing

During testing, the model was presented with a single input form and the Euclidean distance between its output activation and the target output meaning was computed. For *individuation*, each trial was considered a success only if the network's output activation pattern was closest to the target output. For *categorisation*, the model was judged to have correctly identified the category of the referent of the input word if the network's output was closest to a pattern of the same category as the target output.

The simulation was run 40 times per condition, each run using different starting weights, different output category patterns, and a different random assignment of forms to meanings.

Data Analysis

As a test of the ability to replicate and extend Monaghan et al. (2011)'s findings, I first present a brief analysis of the output of our model on only the two conditions from their original paper. For this, the model's performance at the end of each of the four testing blocks was analysed using a repeated measures analysis of variance. I performed a separated ANOVA for each of our dependent variables (individuation and categorisation) with systematicity (arbitrary vs. systematic) as a between subjects factor and experimental block as a within subjects factor.

3.3.2 Replication Results

The results of our model generally align well with the findings of Monaghan et al. (2011)'s original model presented as simulation 1 in their paper (see Figure 3.04). As with their model I find that for Individuation the model learns systematic languages ($M = 0.468$, $SE = 0.013$) better than it does arbitrary ones ($M = 0.267$, $SE = 0.013$; $F(1,319) = 124.5$, $p < 0.001$), although overall our model performs better at the individuation task than does Monaghan et al.'s (Figure 3.04-Bottom), despite using the same learning rate. There was also a significant main effect of block ($F(3,319) = 249.07$, $p < 0.001$) and a significant interaction of systematicity \times block ($F(3,319) = 9.12$, $p < 0.001$).

For categorisation the performance of our model matches that of Monaghan et al.'s, with perfect categorisation from the beginning for the systematic language, which is significantly better than categorisation for arbitrary languages ($M=0.665$, $SE=0.007$; $F(1,319)=1170.05$, $p<0.001$; Figure 3.04-Top). There was also a significant effect of block ($F(3,319)=29.62$, $p<0.001$) and a significant interaction of block x systematicity ($F(3,319)=29.62$, $p<0.001$).

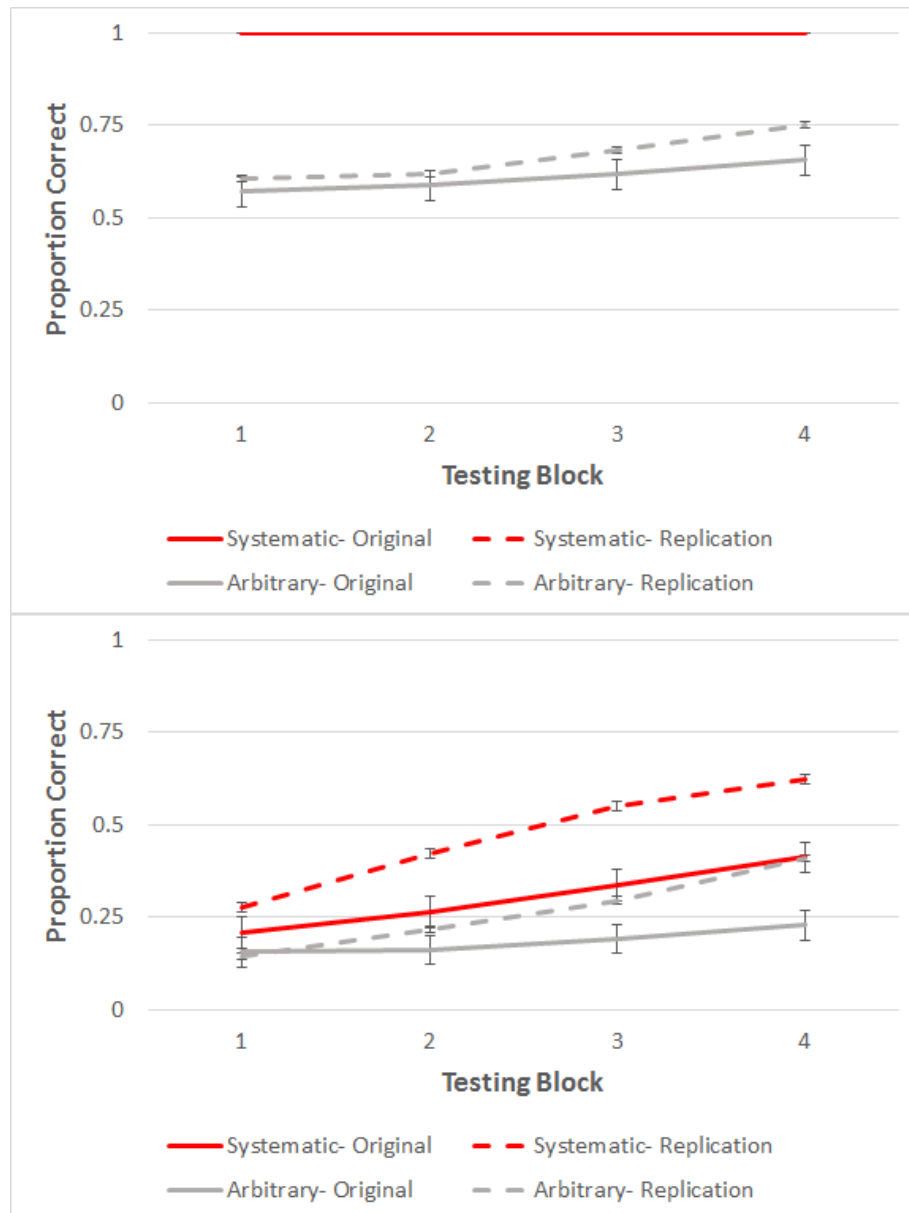


Figure 3.04- Comparison of results between Monaghan et al. (2011)'s original published model and our attempted replication of the model. The top graph shows categorisation, while the bottom graph shows individuation performance. Error bars represent standard error.

3.3.3 Replication Discussion

Our instantiation of Monaghan et al.'s model produced results that were nearly identical for categorisation, but were significantly higher for individuation, despite using the same learning rate. Despite the difference in the ability of the two models to individuate, the fact that the pattern of results between the two versions was nearly identical satisfied our standards for replication.

Individuation and Categorisation

In chapter 2, I hinted at a potential problem with the categorisation and individuation metrics used by Monaghan et al. (2011) in their model and their experiment. Although the metrics measure what they are reported to, their overall usefulness in separating learning into two broad types is actually limited. The central limitation of these two metrics is that they are taken from a single response (which meaning is selected for a given word), and thus not independent from one another. If a run of the model was the individuate perfectly (always choose the correct meaning for a given word), then by the categorisation metric used by Monaghan et al. we would also say that the model categorised perfectly, but is this a fair suggestion? Given this possibility, it would actually be impossible to determine the difference between a model (or experimental participant) that had learned to individuate perfectly and also recognized the structure of the underlying categories

and one that had learned only individual words but nothing about the systematic structure of the word-meaning associations.

To distill the effect that different types of languages have on categorisation then, we require a metric that does not include individuation responses as part of categorisation. To that end, I included a new metric of categorisation that included only cases where the model was unable to individuate correctly, but still chose a token from the appropriate category. In the subsequent analyses, I refer to this metric as *categorisation error*. The inclusion of this metric also allows a simple way to look at the model's improvement in categorisation over experimental blocks. A model that shows a perfect categorisation score using Monaghan et al.'s categorisation metric can do so with any combination of correctly individuating and making categorisation errors, but the two cannot be teased apart easily for analysis. This is actually the case in the original data published by Monaghan et al. (2011) for Experiment 1: systematic languages are categorised at ceiling immediately, but individuation climbs over the course of experimental blocks. Categorisation error gives us a single value that captures this fact and is amenable to direct statistical analysis.

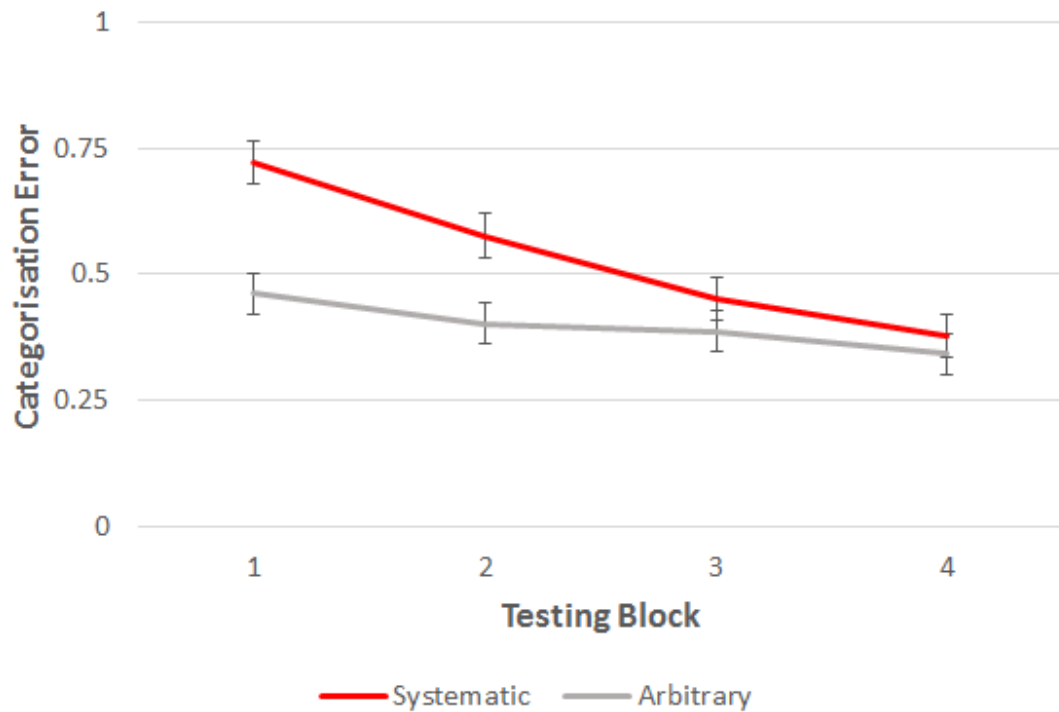


Figure 3.05- A graph of categorisation error from our replication of Monaghan et al. (2011). Error bars show standard error.

Given the graph above, it is easy to see that initially the model has learned individual words poorly, and is selecting meanings that are incorrect, but of the correct type. However, over the course of experimental blocks the model makes fewer categorisation errors. A model that failed to learn individual words at all, however, would see no decrease in its number of categorisation errors over the course of multiple rounds of training and testing.

3.3.4 Methods II

The performance of our full model at the end of each of four testing blocks was analysed using a repeated measures analysis of variance with systematicity (arbitrary vs. systematic), phonological dispersion (clustered vs. dispersed), and phoneme inventory (Monaghan vs. Nielsen) as between subjects factors and testing block as a within subjects factor. I performed three separate rmANOVAs – one for each of our dependent variables (individuation, categorisation, and categorisation error).

Given that the model represents words as sets of phoneme vectors, it is possible to make predictions based on both the contrastiveness of sets of phonemes and the contrastiveness (average Euclidean distance) of the sets of words used in each experimental condition. Broadly, within-class contrastiveness should be correlated with individuation performance (more contrastive = better individuation) and between-class contrastiveness should be correlated with categorisation performance (more contrastive = better categorisation).

Phoneme Contrastiveness

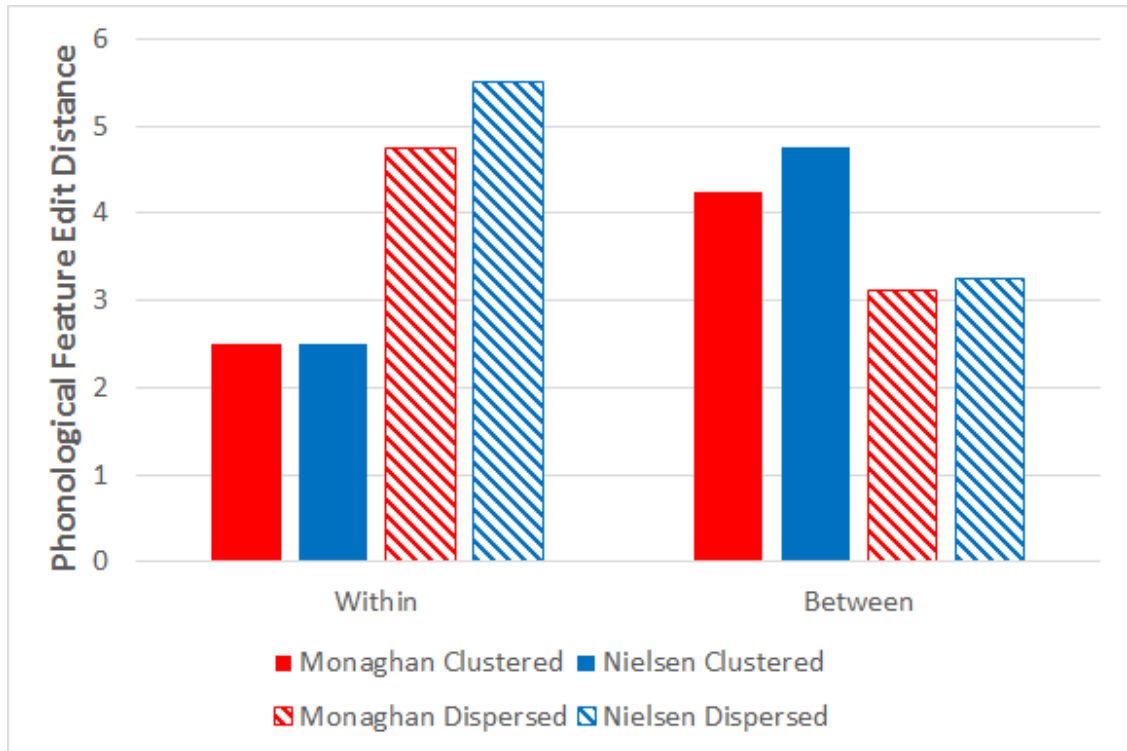


Figure 3.06- The average phonological feature edit distance both within and between the groups of phonemes used to create language for Simulation 1.

Despite the fact that the model only sees complete words, it is also possible to make predictions about performance based only on phoneme inventories: the degree that the contrastiveness of phoneme inventories predicts performance might suggest that the systematicity of entire words is redundant. The phonemes used in languages created from the Nielsen and Monaghan phoneme inventories do not differ in the within-class contrastiveness, and thus should be individuated equally well, while Nielsen's phonemes are slightly more contrastive between classes and thus should be

easier to categorise.

The sets of phonemes used in phonologically dispersed languages are substantially more contrastive within-class and thus phonologically dispersed languages should be individuated more easily than phonologically clustered ones. The increased contrastiveness within these phoneme sets comes at a cost to their between-class contrastiveness, suggesting that phonologically dispersed languages should be more difficult to categorise.

Phonological Feature Contrastiveness

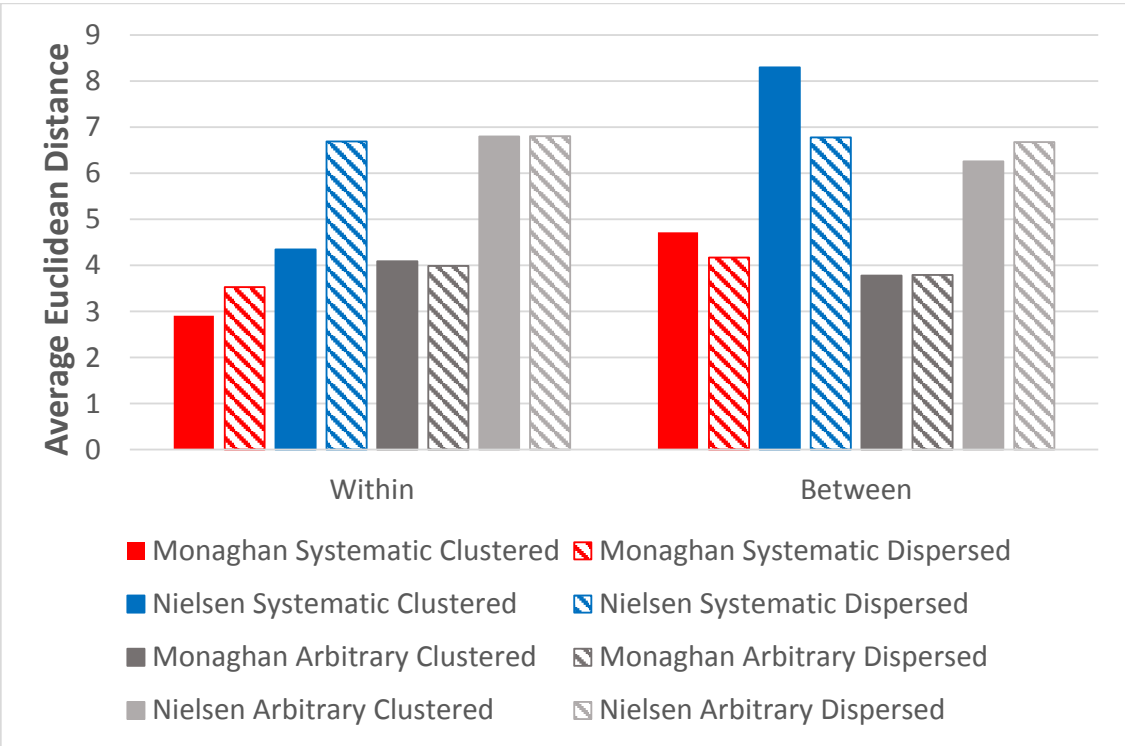


Figure 3.07- The average euclidean distance between phonological feature representations of words use in Simulation 1.

The average euclidean distances of words used in the 8 languages learned by the model make similar predictions to their underlying phoneme composition. Because the words created using the Nielsen phoneme inventory are longer, they are more contrastive both within and between types and should thus be easier to both individuate and categorise. Systematic languages have higher between-class contrastiveness and thus should be easier to categorise than arbitrary languages, which have higher within-class contrastiveness and thus should be easier to individuate. Finally, phonologically dispersed languages have higher within-class contrastiveness and should be easier to individuate than their phonologically clustered counterparts.

3.3.5 Results

For individuation I found significant main effects of systematicity (Arbitrary $M = 0.413$, Systematic $M = 0.567$; $F(1, 1279) = 254.02$, $p < 0.001$), phonological dispersion (Clustered $M = 0.354$, Dispersed $M = 0.62$; $F(1, 1279) = 787.75$, $p < 0.001$), phoneme set (Nielsen $M = 0.532$, Monaghan $M = 0.497$; $F(1, 1279) = 78.51$, $p < 0.001$), and experimental block ($F(3, 1279) = 1406.5$, $p < 0.001$). In addition to these main effects, all two-way and three way interactions were significant (all $p < 0.05$), as was the four-way interaction of all factors ($F(3, 1279) = 3.48$, $p = 0.016$).

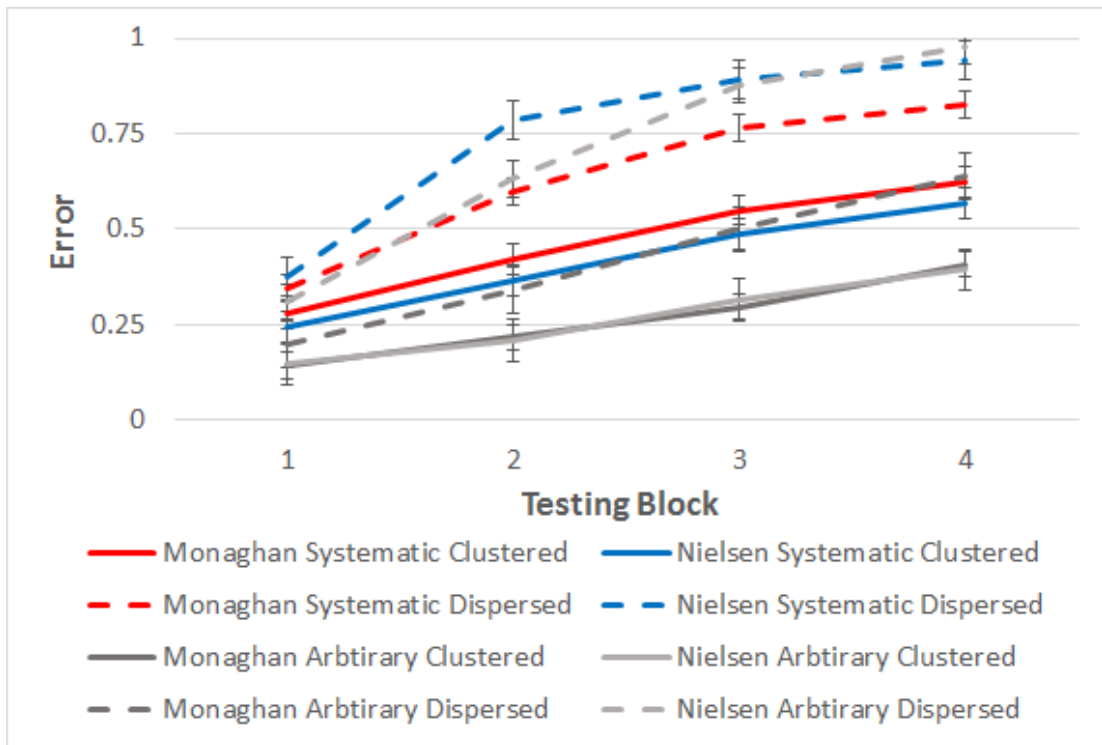


Figure 3.08- Results of Individuation performance for simulation 1. Error bars show standard error.

For categorisation all main effects and their interactions were significant ($p < 0.05$). I found significant main effects of systematicity (Arbitrary $M = 0.743$, Systematic $M = 0.99$), phonological dispersion (Clustered $M = 0.825$, Dispersed $M = 0.908$), and of phoneme set (Nielsen $M = 0.880$, Monaghan $M = 0.853$; Figure 3.09).

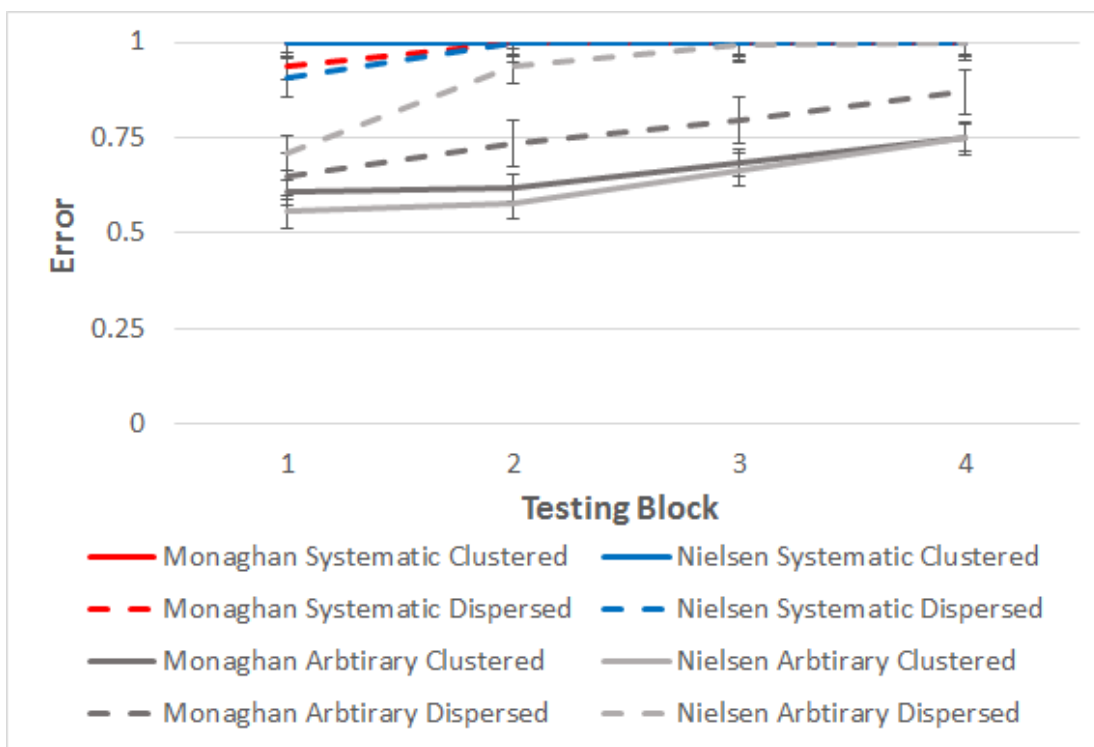


Figure 3.09- Results of categorization performance for simulation 1. Error bars show standard error.

Finally, for categorisation error I again found that all main effects and their interactions were significant ($p < 0.05$). I found significant main effects of systematicity (Arbitrary $M = 0.330$, Systematic $M = 0.4237$), phonological dispersion (Clustered $M = 0.471$, Dispersed $M = 0.283$), and of phoneme set (Nielsen $M = 0.348$, Monaghan $M = 0.406$; Figure 3.10).

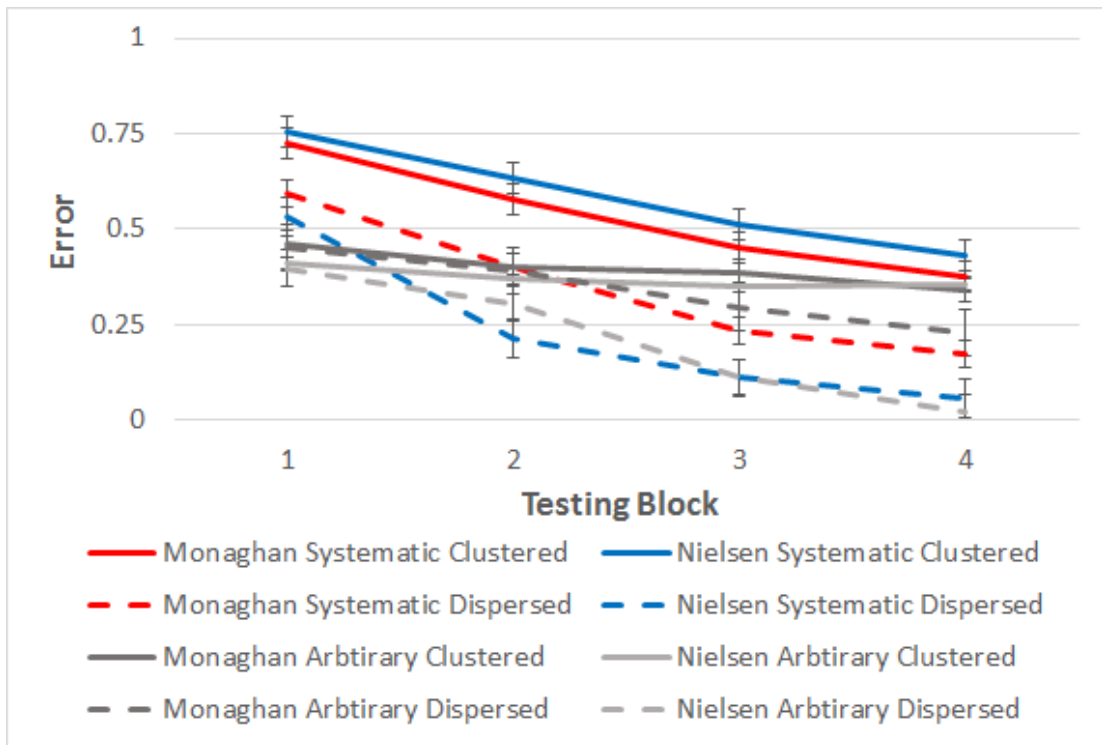


Figure 3.10- Results of categorisation error for simulation 1. Error bars show standard error.

3.3.6 Discussion

The results of our model replicate the findings of Monaghan et al. (2011)’s first model: languages that are systematically structured are learned more easily by the model, both in terms of individuation (associating a form with its specific reference) and categorisation (associating a form with any referent of the correct class). The advantage for systematic languages shows up early in the model, where even at the first testing block the language is already categorizing at ceiling. In line with Monaghan et al. (2011), I suggest that the categorisation benefit account for much

of the difference in individuation between systematic and arbitrary languages. The inclusion of our third metric- that of categorisation error (when the model chooses a meaning of the appropriate type, but not the correct meaning) is further suggestive that this is where the benefits accrue for systematic lexica. We can see that for systematic languages, whenever the model is wrong early on, it is always at least making in-class, rather than out-of-class errors. For arbitrary languages however the choice of a meaning for a given signal does not ensure that other nearby signals will be associated with similar meanings.

I also found a significant effect of phoneme set: labels constructed from the set of plosive vs. sonorant (Nielsen) phonemes were easier for the model to learn in all of their permutations. This second set of phonemes has advantages of its own- the differences between plosive and sonorant phonemes are greater than the equivalent differences between plosive and fricative phonemes used by Monaghan et al. (2011). Additionally, because this second phoneme set is used to construct labels that are twice as long (66 phonological feature units, rather than 33) the model has more features over which to individuate and more reinforcement for categorisation.

Phonologically dispersed systematic languages were learned better than their phonologically clustered counterparts were. The manipulation of phonological dispersion creates systematic languages where the within-class difference for both types of words is maximized, and this predicted that those languages would be

individuated better. The results of this set of simulations thus give clear predictions that can, as in Monaghan et al. (2011), be tested against the learning abilities of human participants.

3.4 Experiment 4

Following Monaghan et al. (2011), I wanted to compare the ability of our model to learn our artificial languages to the ability of human experimental participants. The results of our extension of Monaghan et al.’s model to explore the effect of phonological dispersion on learning suggested that phonologically dispersed languages were significantly easier for the model to individuate, and that systematic phonologically dispersed languages gained the benefit from systematicity (early categorization near ceiling) without the incumbent penalty for individuation of their phonologically clustered counterparts. The fact that the model individuates phonologically dispersed languages better is predicted by both the edit distance of the phonemes used in their construction and the average Euclidean distance of the actual words used: phonologically dispersed languages have higher within-class contrast than do their phonologically clustered counterparts, and are also individuated substantially better.

Here, I evaluate whether human participants produce the same general results as the model: i.e. whether they are sensitive to the phonological feature representations of the words that they use, or whether some other metric of

similarity between words within and between classes is better able to predict human behavior. From the perspective of phonological features, phonologically dispersed languages have greater within-class contrastiveness and should thus be easier for participants to learn. At the level of individual phonemes, however, phonological clustering has no effect on the contrastiveness of words: all words within one type use a single set of phonemes (though the phonemes are not related) and none of those phonemes are used for the creation of both types of words.

As in Chapter 2, I use a signal detection paradigm, compared to the alternative-forced-choice task presented to the model. The rationale for using this experimental methodology, to rehearse from Chapter 2, is that it allows for a separation of pressures for individuation and categorization that are less straightforward given the model's AFC task. The fact that experimental participants respond to three separate types of trials (targets, in-class distractors, and out-of-class distractors) allows for an evaluation of how well they are able to individuate (the difference between performance on target trials and in-class distractor trials) and also their relative certainty about the category structure (out-of-class distractor trials).

3.4.1 Methods

Participants

Participants were 40 students (24 female) and members of the public recruited from the SAGE recruiting service at the University of Edinburgh (mean age 22.15 years). Ethical approval was obtained from the University of Edinburgh in line with BPS guidelines, and informed consent was obtained from all experimental participants. All participants were proficient speakers of English and had normal hearing and normal or corrected-to-normal vision and were paid £2 for their participation, which took approximately 15 minutes.

Participants (n=40) were assigned randomly to each of 4 experimental conditions such that each condition had 10 participants. Conditions 1 and 2 (systematic languages) had further random assignment of participants to subconditions (n=5) that counterbalanced the nature of systematic word-meaning associations. We found no differences between participants in these subconditions (e.g. it did not matter whether animals were paired with plosive or sonorant words), so those subconditions were collapsed for further analysis.

Experimental Design

The experiment conducted here used a signal detection protocol to measure the ability of participants to learn associations between novel labels and their meanings in a 2 (systematic vs. arbitrary) \times 2 (phonologically clustered vs. dispersed) factorial design. For the *systematic clustered* condition, where signal meaning mappings were phonologically systematic, there was a correspondence between the category of the picture and the phonology of the labels (e.g. all plosives were paired with animals), whereas for the *systematic dispersed* condition the mappings between form and meaning were systematic but not based on phonology. The remaining two conditions, *arbitrary clustered* and *arbitrary dispersed*, were expected to be equivalent, as the phonological characteristics of a label were not predictive of the category of its meaning in either case.

Label Stimuli

The word stimuli used for the experiment here were generated using the “Nielsen” phoneme set used above in Simulation 1 but instead of using a fixed set of 6 of the possible 8 generated words, each participant saw a random sample of 6 of the 8 possible words. The stimuli were created identically to those in Experiment 3 of Chapter 1- thus, each syllable was produced by Apple talk and then during the

experiment those syllables were presented sequentially to form words that were unstressed.

		Clustered		Dispersed	
		Type 1	Type 2	Type 1	Type 2
Systematic		mo mo mo	ti ti ti	mu mu mu	no no no
		mo mo nu	ti ti ke	mu mu te	no no ki
		nu nu nu	ke ke ke	mu te mu	ki no no
		nu nu mo	ke ke ti	te te te	ki no ki
		nu mo nu	ti ke ke	te te mu	no ki ki
		mo nu mo	ke ti ti	te mu te	no ki no
Arbitrary		mo mo mo	ti ti ti	mo mo mo	nu nu nu
		nu nu mo	ke ke ke	mo te me	ki nuki
		mo nu mo	ke ti ti	te mo te	nuki ki
		ke ke ti	nu nu nu	nu nuki	mo mo te
		ti ke ke	nu mo nu	ki nu nu	te te te
		ti ti ke	mo mo nu	nuki nu	te te mo

Table 3.04- A sample of possible languages for the four conditions of Experiment 4

Image stimuli

Word meanings were selected from two distinct categories: animals and vehicles, and were taken from a variety of online sources using Google Image search; images were extracted from their background and placed on a white background, then standardized for size and resolution (Figure 3.11).

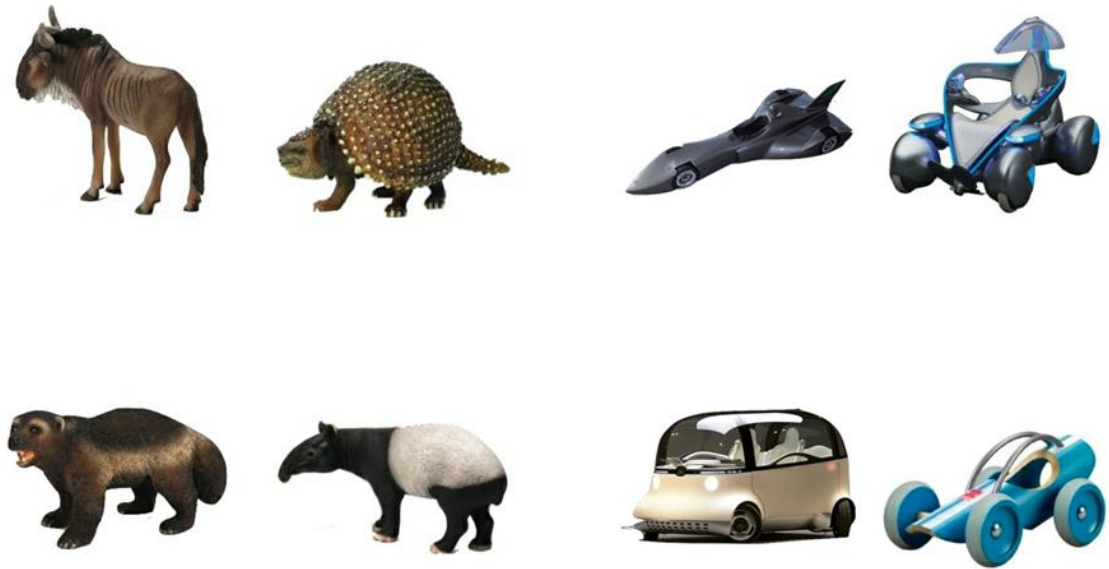


Figure 3.11- A sample of the images used in Experiment 4.

Procedure

Familiarisation

Prior to training, participants were exposed to all of the labels used for their version of the experiment, absent their referents, via headphones in two randomized blocks, to familiarize them with what the novel labels sounded like.

Participants were then instructed that they would be presented with pairs of labels and images and it was their task to remember the pairs of images and labels

that they saw. They were informed that after training they would be tested on their ability to recognize label-meaning pairs that they had previously been exposed to.

Training

After familiarization, participants were shown proper training trials, where an image was shown on screen for 750 milliseconds before a label was played to them via headphones, twice, with a one second break between each presentation. One second after the second presentation of the label, the image would disappear and a new training trial would begin. During training, each label-meaning pairing was presented twice in two blocks whose order was randomized.

Testing

At test, participants were presented with pairs of labels and images. Each label and image was seen a total of four times across three types of experimental trial, for a total of 48 trials. One quarter of the trials (12) were *target* trials, where the presented label-image pair was one to which the participant had been exposed during training. One quarter of the trials (12) were *in-class distractor* trials, where the presented label-image pair was not one that had previously been learned but where the image was of the same type (animal or vehicle) as the image originally presented with the label (e.g. if ‘*munomu*’ was a label for a car, it might be presented with another vehicle as an in-class distractor). The remaining trials (24) were *out-of-class distractor* trials,

where the label was presented with an image of the opposite type as the one it had originally been paired with. On each trial participants responded via keyboard, pressing either the ‘z’ key for “no” or the ‘/?’ key for “yes” on a given trial. After their selection the experiment proceeded to the next trial. The experiment was conducted using an interface created with Livecode v 5.02.

Contrastiveness

I calculated the contrastiveness of the possible languages in each of the four experimental conditions using two metrics. First, I used the phonological features of possible sets of words for each condition to calculate an average Euclidean distance within each type of words and between each type. Second, I calculated the average Hamming distance of possible sets of words for each condition based on their phonemes. The values of those contrastiveness calculations for each condition can be seen below in Figure 3.12.

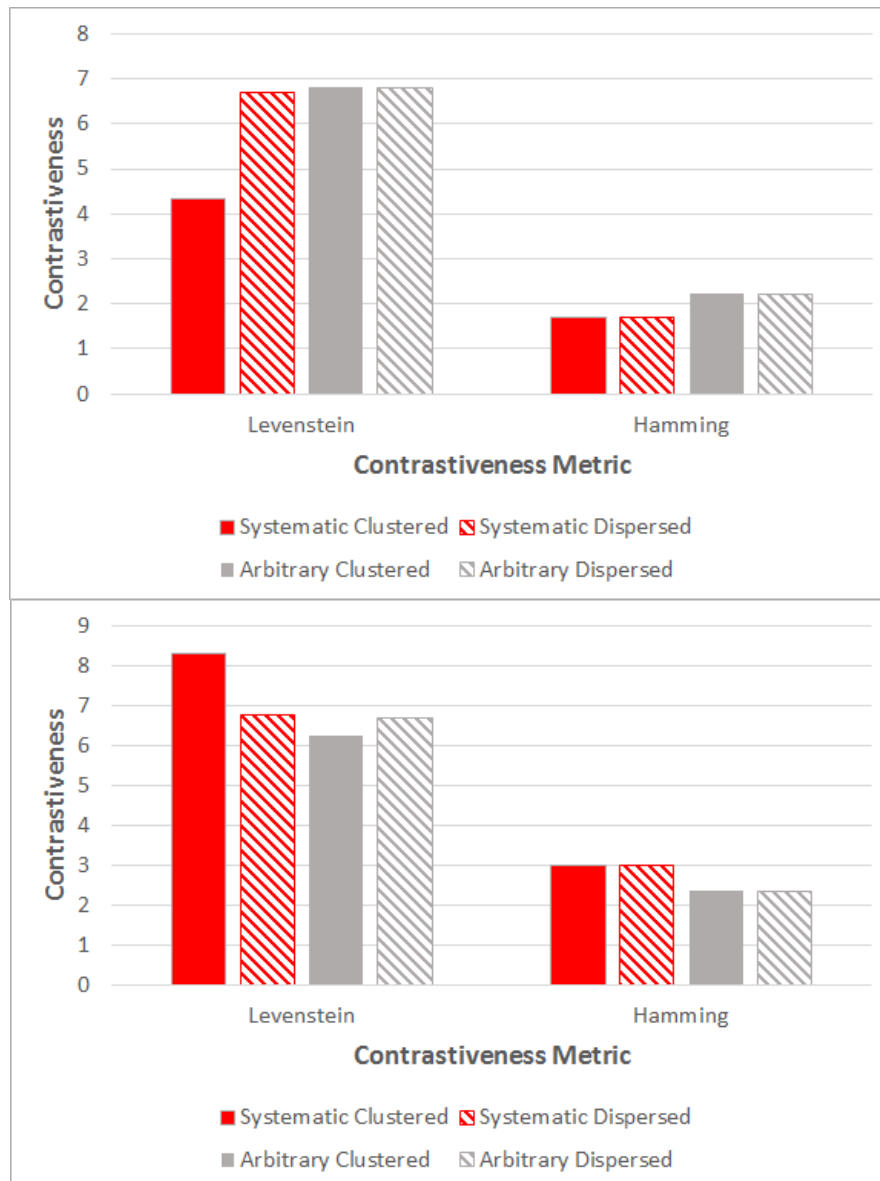


Figure 3.12- The average within-class (top) and between class (bottom) contrastiveness of lexica for each of the four conditions of Experiment 4, calculated using the phonological feature euclidean distance (left) and the simple Hamming edit distance (right).

Both metrics suggest that systematic lexica have lower within-class contrastiveness than their arbitrary counterparts, so should be harder to individuate. Both

contrastiveness metrics predict no difference in learnability between arbitrary lexica, so both suggest that arbitrary lexica, whether they are phonologically dispersed or phonologically clustered, should be equally learnable.

The euclidean distance based on phonological features predicts an effect of phonological clustering- phonologically dispersed languages have higher within-class contrastiveness than their phonologically clustered counterparts. Hamming distance based on phonemes however does not predict this effect: the distance within groups based on their Hamming distance is the same regardless of whether or not the words are composed from similar phonemes or not.

Data Analysis

As with the results of Experiments 1-3 from Chapter 2, responses were scored according to a signal detection paradigm; on target trials “yes” responses were scored as hits, with “no” responses as misses, while on distractor trials of both types “yes” responses were scored as false alarms with “no” responses scored as correct rejections. These responses were transformed to a d' value for each participant, which I used as an omnibus measure for a general linear model with systematicity (arbitrary vs. systematic) and phonological clustering (clustering vs. dispersed) as factors. Additionally, I conducted a repeated measures analysis of variance with systematicity and phonological clustering as between subjects factors and trial type (target vs. in-class distractor vs. out-of-class distractor) as a within-subjects factor.

3.4.2 Results

Signal Detection

For the omnibus test of performance using d' I found a main effect of systematicity, in line with predictions based on both of my contrastiveness metrics (and previous work): Systematic languages ($M = 1.41$) were easier to learn than Arbitrary ones ($M = 0.503$, $F(1) = 16.12$, $p < 0.001$). I did not however find a significant effect of phonological clustering (Clustered $M = 1.09$, Dispersed $M = 0.81$; $F(1) = 1.63$, $p = 0.21$), and only a marginal interaction of the two ($F(1) = 3.87$, $p = 0.057$), supporting the predictions of hamming distance as a contrastiveness metric; see Figure 3.13.

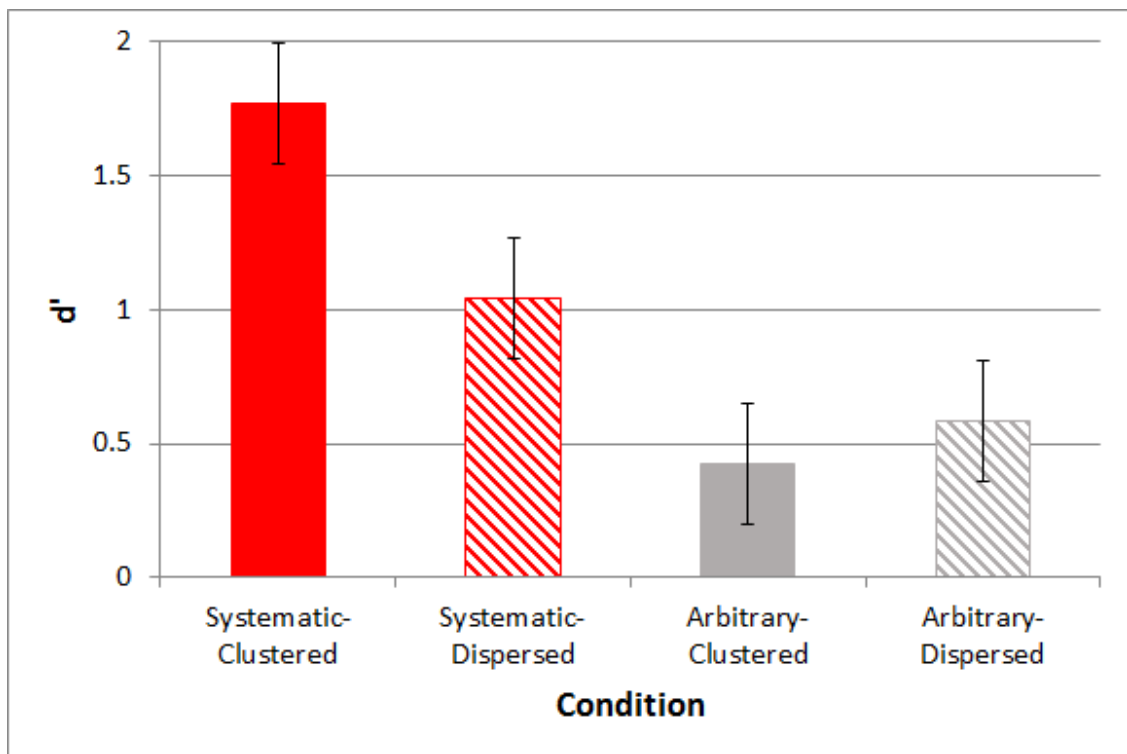


Figure 3.13- Omnibus d' results from Experiment 4 demonstrate a significant main effect of systematicity: systematic languages are easier to learn in this task than arbitrary ones. Error bars represent standard error.

Repeated measures Analysis of Variance

The repeated measured analysis of variance revealed a significant main effect of condition: systematic language learners performed significantly better than arbitrary language learners ($M = F(1,119) = 7.55, p = 0.009$). I found no significant effect of phonological dispersion ($F(1,119) = 0.55, p = 0.463$) and no significant interaction of systematicity x phonological dispersion ($F(1,119) = 2.03, p = 0.163$). There was a significant main effect of trial type ($F(2,119) = 13.68, p < 0.001$): post hoc

analysis using the Tukey-Kramer Multiple comparison test showed that participants performed significantly worse on in-class-distractor trials ($M=0.52$, $SE= 0.0312$) than on either Target ($M= 0.68$, $SE=0.0312$) or out-of-class distractor trials($M= 0.72$, $SE=0.0221$).

In addition to main effects of systematicity and trial type, I found a significant interaction between trial type and systematicity ($F(2,119)=11.44$, $p<0.001$). I found, using post-hoc analysis, that participants who learned arbitrary languages did not perform significantly differently on the three trial types (Target: $M= 0.60$, $SE= 0.035$; In-Class Distractor: $M= 0.58$, $SE= 0.035$; Out-of-class Distractor $M=0.59$, $SE= 0.025$; $F(2,59)=0.11$, $p=0.89$). Systematic language learners however performed significantly different depending on trial type ($F(2,59)= 17.48$, $p<0.001$): Systematic language learners performed worst on in-class distractor trials ($M=0.454$, $SE= 0.052$) than they did on either target ($M=0.754$, $SE= 0.052$) or out-of-class distractor trials ($M=0.85$, $SE= 0.036$), on which they performed equally well (Figure 3.14). There was no significant three-way interaction of systematicity x phonological dispersion x trial type ($F(2,119)= 2.18$, $p= 0.12$).

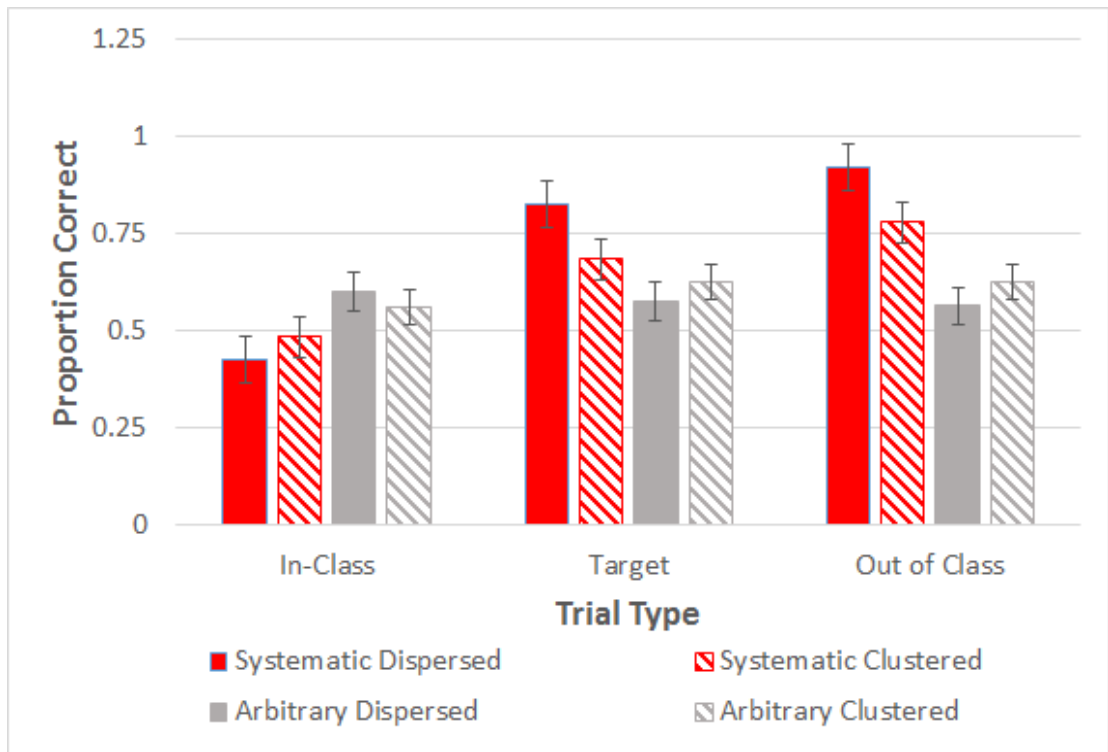


Figure 3.14- *Effect of the interaction of experimental condition and trial type on the proportion of correct responses. Participants in the arbitrary language condition performed equally well regardless of the type of experimental trial, while systematic language learners performed best on out-of-class distractor trials and worst on in-class-distractor trials.*

Finally, I compared performance on each trial type between the two conditions using further rmANOVAs. I found that systematic language learners performed systematically better on target trials ($F(1,39)= 7.22, p=0.011$) and out-of-class distractor trials ($F(1,39)= 29.78, p<0.001$) than did arbitrary language learners. However, arbitrary language learners performed only marginally better on in-class distractor trials ($F(1,39)= 3.10, p=0.086$) than systematic language learners.

3.4.3 Discussion

The results of this experiment support the general findings of Monaghan et al. (2011) as well as the findings of the three experiments presented in Chapter 2 of this dissertation: systematic languages were easier to learn for both our replication of Monaghan et al.'s model and my experimental participants. Specifically, in both cases, systematic languages aid in the task of categorization, while arbitrary languages, by virtue of being more contrastive, aid the task of individuation.

Based on the results of Simulation 1, I expected that human participants would be sensitive to the phonological features of the languages used in Experiment 4: phonologically dispersed languages would create a benefit to categorization but also be easier for the experimental participants to individuate. Calculating contrastiveness based on the phonological feature mapping from Harm & Seidenberg, by which words are represented in the simulation, predicts that there should be an effect of phonological clustering, whereas calculating contrastiveness based on Hamming distance (as in Chapter 2) predicts no difference in learnability between phonologically clustered and phonologically dispersed languages. The fact that human learners performed no differently on the two trial types suggests that Hamming distance at the level of phonemes is the more appropriate of the two metrics for predicting human learning.

3.5 General Discussion and Conclusions

The results of Chapter 2 showed that the idea of contrastiveness is central to determining whether languages can be systematically marked in such a way to produce a benefit to categorization learning without inducing a concomitant penalty to individuation. Because contrastiveness is central to the tension between a positive learnability pressure favoring systematicity and a negative pressure favoring arbitrariness, it is crucial to understand what features are relevant to human language learners. Monaghan et al. (2011), by creating a simulated neural network that used phonological features to represent words, and demonstrating that the model performed similarly to human language learners, suggested that phonological features are most relevant to human perception.

However, both the phonological features and an analysis considering only the phonemes used both make the same predictions about the learnability of systematic vs. arbitrary languages. Based on the phonological feature edit distance within and between classes of words, systematic mappings between words and meanings reduce the within-class contrastiveness of words while making the between-class distance larger. However, considering phonemes as discrete units leads to the same predictions. To test which of these two possibilities was a better model for human perception, I introduced an additional factor of phonological clustering on which

both our replication of Monaghan et al.'s model and my experimental participants were tested.

The results of our model suggested that phonologically dispersed languages were easier to learn, especially using our individuation metric. Because /f/ and /g/ are more different than /g/ and /k/ based on their phonological feature representations, mapping /f/ and /g/ systematically to one type of meanings allowed the model to learn about category information without leading to confusion of individual words to each other.

The results of my human language learners, however, did not support the findings of the model: for human language learners there was no effect of phonological clustering, suggesting that the systematic use of phonemes for marking category structures results in benefits for categorisation learning and penalties for individuation learning regardless of how similar phonemes within a category are to one another. Human participants seem to be able to learn systematic languages that are not based on phonological clustering- i.e. they are able to map dissimilar phonemes onto nonetheless similar meanings. Specifically, participants learned that pairs of features that are unrelated to each other can nonetheless be predictive of the same category, despite phonologically-similar features being used as labels for an opposing category. Specifically, because the experimental participants we typically employ are already familiar with a language, the differences between say /p/ and /t/

may be more salient than they are when represented as phonological features to a language naïve model. There is at least one interpretation that is more favourable towards the phonological feature: children might learn more similarly to Monaghan et al.'s original models, but whether this is true is an open question. A second interpretation is that phonological feature representations are unable to capture the perceptual salience of labels for this kind of task at all.

One interesting possibility, first raised by Gasser (2004), is that the saturation of the available signal space will have important implications for whether arbitrary or systematic languages are ultimately easier to learn. That is, systematic language are easier to learn when there is sufficient space for there to be systematic associations without those labels actually becoming too similar to one another, and thus confusable. With the inclusion of increasingly large numbers of labels (which increases the saturation of the available space) the benefit for systematic languages becomes increasingly small and eventually inverts. Using our experimental protocol, it's possible that given a sufficiently large number of stimuli per image type learning only category markers without any ability to individuate tokens would be an optimal strategy (given a memory limitation for similar tokens). The balance between systematicity and arbitrariness is likely to not only be based on the design of the experimental task (or the typical situations under which language learning is conducted) but also the overall similarity of labels to one another.

The results of the computational models and artificial language learning experiment presented here provide further evidence that systematic languages, where there are system-level associations between form and meaning, allow language learners to learn to categorise novel words more easily. In contrast to previous work however I found additional evidence that at least in some cases systematic languages are also easier to individuate- this is especially likely to be the case when lexica are small (Gasser, 2004). This might have important implications for the trajectory of language learning: early in acquisition the size of lexica will of necessity be small, so preferentially teaching systematic labels to new learners would be optimal. Ultimately, as the lexica of these learners grow, moving to more arbitrary signal-meaning mappings would become easier. This possibility is supported by recent findings which suggest that early acquired words are more likely to be systematic than late acquired ones (Monaghan et al., 2014), suggesting that although overall the lexicon does not appear to be systematic, the age at which words are acquired may reflect the fundamental division of labour between systematicity and distinctiveness.

One additional benefit that systematic languages might have over arbitrary ones that is not tested in the experiment or simulations is that systematic languages should allow for generalisation to entirely novel tokens. In this case, new signal-meaning mappings that are congruent with previously learned systematic relationships might be accepted at rates above chance. Although I have not directly explored that possibility here, it seems to be one that would be rather straightforward

to test, using either artificial language stimuli or real-world systematic form-meaning associations like those found in English phonaestheme clusters. Thus, when presented with an unfamiliar word like ‘*gloom*’ it might be easier for a new participant to learn that it was a word having to do with light. Here we might see a similar division of labour between arbitrariness and systematicity: systematicity might make predicting meanings given a new signal or creating sensible signals for a new meaning easier, but it might also cause overgeneralisation so that non-systematic labels cannot be learned as easily. This division of labor might again be productive, with arbitrary form-meaning mappings actually allowing systematic relationships to persist without collapsing into a signal meaning.

The work presented here has additionally focused on systematic signal-meaning mappings that are conventional rather than motivated. The associations that I have tested are isomorphic in the sense that specific features of signals are mapped onto features of meanings, but the direction of these mappings can go either way. For example, my data do not suggest that plosive consonants are any better when paired with animals than with vehicles, but there are a number of associations between phonological features and object characteristics that are motivated by the perceptual or cognitive apparatus of the language learner, like the Bouba-Kiki effect (Kohler, 1929; Mauer et al., 2006; Nielsen & Rendall, 2011, 2012, 2013), where voiceless plosive consonants are associated with jagged image forms and voiced sonorants with curved image forms. It is possible that motivated signal-meaning

associations of this type are also important for language learning, as they allow for generalisation without previous experience with the language (this could be thought of as generalisation from a perceptual prior, rather than a learned one).

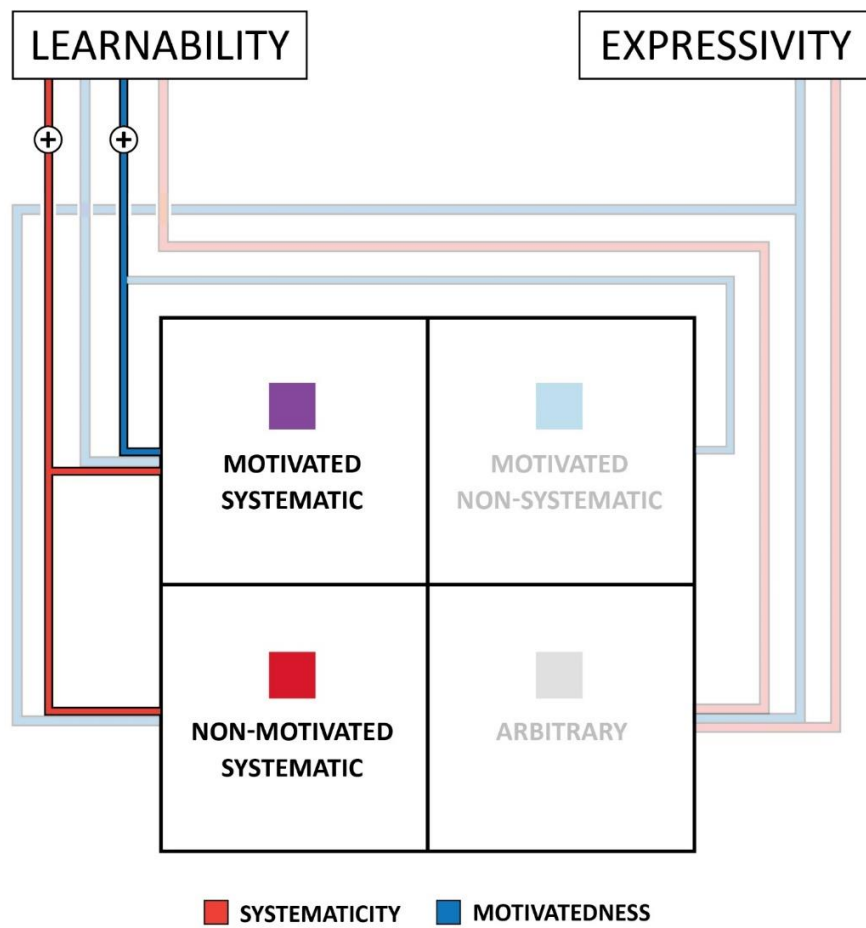
Finally, I suggest that what appears to be increased realism in computational models of human language learning is not always beneficial. The fact that human languages are made up of well-defined phonemes does not mean that the phonological feature representation of those phonemes is necessarily appropriate as a coding for the percept of those languages for human learners.

The degree to which natural languages take advantage of the potential learnability benefits for systematic languages is currently not well understood, with the majority of research focusing on iconicity, rather than systematicity, as a plausible bootstrapping method for language acquisition (Imai et al., 2008; Monaghan, Mattock, & Walker, 2012; Nygaard, Cook, & Namy, 2009). A more complete understanding of how structural regularities at the level of the lexicon can influence language learning will need to take into account differences between motivated and conventional systematic form meaning mappings (Nielsen et al., in prep.) The recognition that the original categorisation metric used by Monaghan et al. (2011) is composed of a combination of correct and incorrect responses to a task that is explicitly about individuation further makes the suggest that systematic languages “aid categorisation” one that is difficult to support with the data. In fact, neither the experimental participants nor the model is ever asked to provide any

responses about what category a word-meaning pair belongs to. The ‘categorisation benefit’ for systematic languages can thus only really be pointed to as an increased probability of guessing the correct label from a learned category. Although I do not test for anything that could be conceptualized as a real categorisation task in this chapter, Chapter 4 introduces an experimental paradigm where participants are asked onto to categorise.

Chapter 4

Motivatedness and categorisation



Foreword:

Declaration of submission for publication

The contents of Chapter 4 represent an article submitted for publication to the *Journal of Experimental Psychology: Language, Memory, and Cognition*. As the chapter is only submitted, but has not been published, it has been modified to the format of the thesis content generally. The body text of this chapter has, however, not been modified from the form in which it was submitted for publication, other than changing the names of the experiments presented here to line up with the experiments in the rest of the dissertation. Because this chapter represents a potential publication, it has been edited collaboratively to a greater degree than other work presented in this dissertation, although the writing and statistics for this Chapter were completed primarily by the author of this dissertation.

Introduction of submission for publication

The text of this chapter is largely congruent with the central narrative of this thesis, but differs slightly in its use of terminology, using terms like *iconic* and *sound-symbolic* more commonly than they are used in the rest of the dissertation. The main thrust of this chapter is an exploration of the learnability advantage for motivated associations between words and meanings, and an attempt to separate that learnability benefit from one related to systematicity. The results shown here neatly demonstrate the early advantage that motivatedness provides to naïve learners while simultaneously demonstrating that conventionally (non-motivated) systematic languages are ultimately equally learnable. The similarity of the learnability between motivated and non-motivated systematic languages provides us with a rationale to use motivated systematic associations in Chapter 5, where we test the effect of decreasing contrastiveness on language learning directly. Because this chapter was produced more collaboratively than other chapters, I make use of the pronoun “we”, rather than “I” throughout

Additionally, although we do not make it explicit in the same way that I do here in Chapter 3, the experiments presented in this manuscript are tests of categorisation, rather than individuation. Thus, the results show that participants are able to match shared features of words onto shared features of meanings.

Introduction

The traditional linguistic assumption of the arbitrariness of the sign (de Saussure, 1983; Hockett, 1960) holds that words and their meanings are related only by linguistic convention - after all, there is nothing ‘dog-like’ about the word dog, and any other label established by local convention could equally well be the word for a ‘dog’. There is, however, increasing evidence for the pervasiveness of *systematic* mappings between words and meanings in natural languages. Systematicity exists in a lexicon when some feature of a set of words can be reliably mapped onto some feature of their meaning; that is, where there is an isomorphism between some dimension of meaning and some dimension of form in the lexicon (e.g. in the phonological form of words). Although it is widely accepted that natural languages are massively systematic above the level of the lexicon (i.e. in their morphosyntax), the idea that the lexicons of natural languages might be systematically organized has only recently begun to be seriously considered. It has been found, for example that phonological features (in the simplest case, length) are predictive of grammatical categories (Farmer, Christiansen, & Monaghan, 2006; Fitneva, Christiansen, & Monaghan, 2009; Kelly, 1992); furthermore, Monaghan et al. (2014) have shown that across the entire lexicon, the English language is more phonologically systematic than would occur by chance – although the effect is not particularly large, features of sound and meaning are mapped onto one another in statistically reliable ways.

These systematic mappings between words and meanings can be seen to be either *motivated* or entirely established by *convention* - a fundamental difference that until recently has been ignored. Motivated systematic mappings involve some degree of iconicity or transparency in the mapping between meanings and signals, and go by a number of names in the literature, being called variously *iconic* (e.g. Ahlner & Zlatev, 2010), *sound-symbolic* (Hinton, Nichols, & Ohala, 1994), *crossmodal* (e.g. Cuskley, 2013), or *phonosemantic* (Akita, 2011). A large number of languages use *mimetics* or *ideophones*, where the words that describe an event or object are somehow imitative of that event or object, for example, in Japanese the word ‘*goro*’ refers to a heavy object rolling, while the word ‘*koro*’, starting with a voiceless consonant, refers to a light object rolling (e.g. Imai et al., 2008). Examples of this type of association can also be seen in English, in onomatopoeic expressions (e.g. ‘*crash*’ is meant to be imitative of the sound that it describes). Ideophones and onomatopoeic expressions fall under the general heading of *sound-symbolism*, and are thus considered motivated associations between word and meaning, driven in part by the perceptuo-cognitive organization of language users (Cuskley, 2013; Nielsen, 2011; Ramachandran and Hubbard, 2001).¹ One consequence of motivated

¹ Although specific mechanisms linking sound and meaning are rarely offered, one intriguing possibility is that the same types of cross-modal associations as those seen in synaesthetes might mediate this process and be reflective of human cognitive organization more generally (Maurer & Mondloch, 2004; Simner, 2006, 2012).

associations between single words and single meanings is that a group of such motivated mappings will exhibit systematicity, since the motivatedness of individual mappings ensures that a group of such mappings exhibits systematicity as defined above: their common semantic features map to a shared dimension of form.

In contrast to these motivated associations², conventional systematic mappings between words and meanings are non-motivated: the observed isomorphism between meaning and form is a function of a particular linguistic convention, rather than the perceptuo-cognitive organization of the language's users. For example, in English the word-initial *gl-* cluster is found in a number of words associated with light and vision (e.g. *glint*, *gleam*, *glare*, *glitter*, etc.; but, note that not all *gl-* words are part of the cluster), but this seems likely to be due to

Specifically, the synaesthetic account of these motivated associations between sound and meaning places them firmly in the realm of being explained by perceptuo-cognitive biases in language processing and/or production, rather than being observed in language users as a function of previous learning via exposure to their language.

² The differences between conventional or motivated connections between words and meanings are not always immediately obvious - in each case that a systematic connection between words and meanings are found we must apply psycholinguistic techniques to determine the degree to which these associations are purely conventional or motivated (Nielsen and Rendall, 2012).

language-specific clustering and history rather than *gl-* being a particularly effective or evocative consonant segment for denoting light (Cuskley and Kirby, 2013).

4.1 Background and Rationale

Monaghan et al. (2014) show that the language-wide tendency for systematicity in English is most pronounced in early-acquired words, suggesting that this systematicity might facilitate learning. Similarly, motivated connections between words and objects provide language learners with *a priori* expectations about the likely meaning of some words, and might therefore facilitate word learning. We review this evidence below. It has furthermore been suggested (Asano et al., 2015; Dingemanse et al., 2015; Lockwood & Dingemanse, 2015; Imai & Kita, 2014; Perniss & Vigliocco, 2014) that the presence of motivated signal-meaning mappings might bootstrap the acquisition of mappings that are not sound symbolically motivated, a possibility which has to date received little empirical support. Imai & Kita (2014) suggest that the presence of sound symbolism allows infants and toddlers to establish lexical reference that can then be extended from motivated forms to purely conventional forms. This bootstrapping hypothesis is currently the most well-developed one in the literature, with the exact nature of how sound-symbolism might influence real-world language learning left unexplained in most other places (e.g. Nielsen, 2011). One possibility is that learning motivated word-meaning correspondences might directly increase the capacity to learn conventional words,

e.g. if processing of sound-symbolic associations requires less time or cognitive resources, this will free up those resources for learning of conventional tokens.

Research using both adults and children has suggested that motivated tokens are easier to learn, even cross linguistically (Imai et al., 2008; Nygaard, Cook, & Namy, 2009). One particularly well-studied motivated form-meaning connection is known as the Bouba-Kiki effect (Kohler, 1929; Maurer et al., 2006; Pexman & Sidhu, 2014; Ramachandran & Hubbard, 2001), where jagged images are associated with words containing plosive consonants and curved images are associated with words containing sonorant consonants (and to a lesser degree, rounded vowels: cf. Fort, Martin & Peperkamp, 2015; Nielsen & Rendall, 2011, 2013). Three recent papers (Aveyard, 2012; Monaghan, Mattock, & Walker, 2012; Nielsen & Rendall, 2012) explored the degree to which such motivated connections between signals and meanings enhanced the learnability of an artificial language. In all three cases, the authors compared the ability of participants to learn associations that were congruent with known sound-symbolic associations (e.g. that a word like '*teka*' should be paired with a jagged image) or incongruent with such associations (e.g. *monu* might be associated with a jagged image; see Figure 4.01).

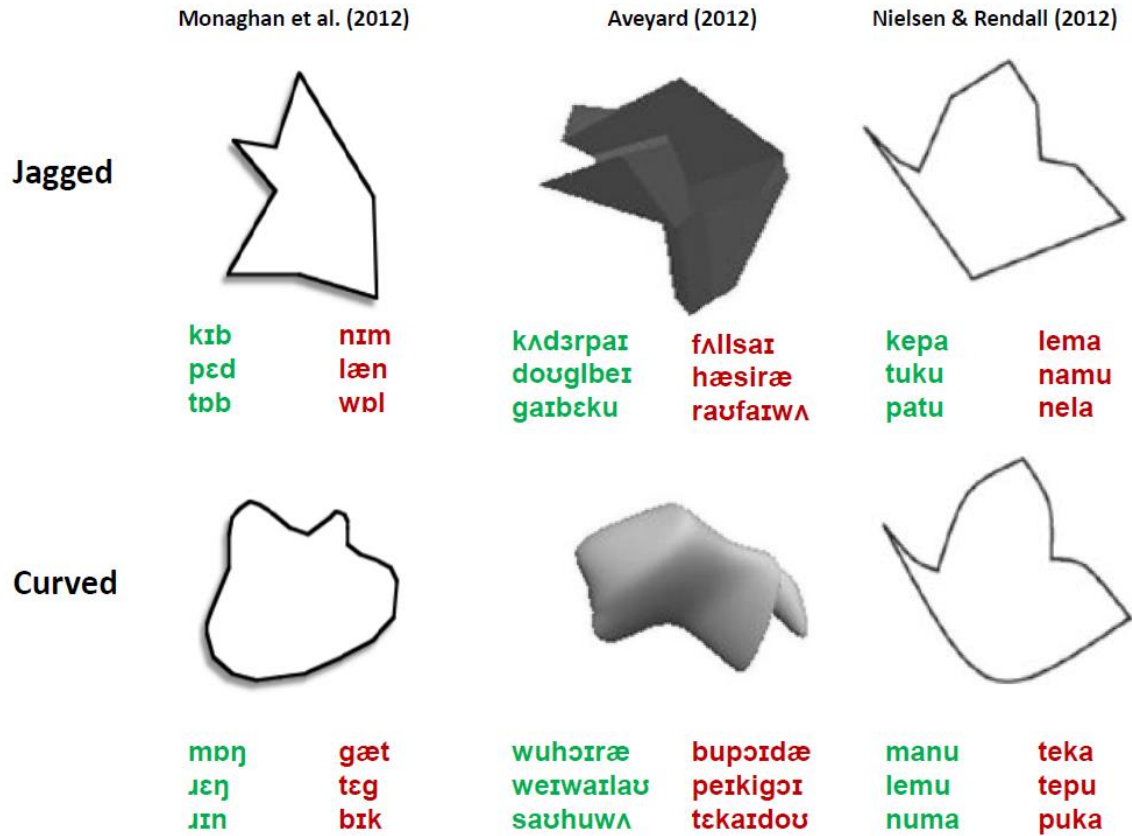


Figure 4.01- Sample stimuli used in previous experiments. Labels in green below each image are congruent with known sound symbolic biases. Labels with red are incongruent with (the opposite of) these biases.

In all three experiments, participants were able to learn congruent associations between words and images better than incongruent associations. This suggests that motivated connections between words and meanings might facilitate language learning.

The artificial languages used in these three studies take advantage of iconic or sound-symbolic associations between phonemes/phonetic features and visual

features of objects. However, in addition to being motivated, these lexicons are also necessarily systematic in the same way as the previously discussed conventional systematic systems like phonaesthemes: even without a bias to associate, say, plosive consonants with jagged images, a learner of one of these languages might nonetheless recognize the presence of the systematic tendencies in the lexicon. A recent study demonstrates that purely conventional (i.e. non-motivated) systematic connections between words and meaning improves learnability of artificial lexicons (Monaghan et al., 2011; but, cf Nielsen et al., in prep). These studies looked at lexicons where some feature of words is mapped onto a feature of a class of meanings (Monaghan et al. use a systematic artificial lexicon where words constructed from plosive consonants refer to objects, and words constructed from fricative consonants refer to actions), and focused on how systematicity influenced the ability of learners to individuate (identify the correct meaning for a given word) and categorize (identify a meaning of the correct category, i.e. object or action). Their results show that systematic associations between words and meanings facilitate the process of categorization, but this increase in category learning comes at a cost to individuation - systematic associations between words and meanings necessarily constrain the available signal space (since all words in a category share many features, i.e. sound alike), making differentiating between words more difficult (Monaghan et al., 2011; Gasser, 2004; but, cf. Nielsen et al., in prep).

Returning to the literature showing a learnability advantage for motivated mappings: by comparing lexicons that are both systematic and motivated to lexicons that are systematic but counter-motivated (i.e. incongruent with known sound-symbolic biases) these studies might be either over- or under-estimating the learnability benefits of motivatedness. Firstly, the systematicity of the counter-motivated lexicons might diminish (or perhaps exacerbate) the cost of being counter-motivated. Secondly, the difference in learnability between motivated and counter-motivated lexicons might reflect either a benefit for motivatedness *or* a penalty for counter-motivatedness, with the comparison of the two extremes making it impossible to tease these two possibilities apart. This might have important implications for natural languages: should we expect natural languages to favor motivated associations between signals and meanings, or avoid counter-motivated associations, or both? Finally, in reference to the bootstrapping hypothesis introduced above, because previous studies entangle motivatedness and systematicity, we cannot be sure of the degree to which learning benefits are a function of motivatedness rather than systematicity: we have evidence both that motivated associations are learned more easily than counter motivated associations, and that conventionally systematic lexicons are (in some cases) learned more readily than purely arbitrary lexicons, but this tells us nothing about whether motivated systematic lexicons and conventional systematic lexicons are learned differently.

4.2 Investigating motivatedness

Here, we report the results of two experiments designed specifically to compare the learnability of systematic lexicons (mapping words to shapes) that vary in their motivatedness, including comparing the learnability of motivated systematic lexicons to purely conventional systematic lexicons.

In Experiment 5, we compare the learnability of artificial lexicons where phonological features like plosivity are mapped to shape features in ways that are either motivated (e.g. plosives map to jagged shapes), counter-motivated (e.g. plosives map to curved shapes), conventionally systematic (e.g. dental fricatives map to jagged shapes, palate-alveolar fricatives map to curved shapes), or partially motivated (e.g. jagged shapes are labeled with a mix of plosives and fricatives; see Table 4.01 below). Experiment 1 shows that the benefit of motivatedness relative to purely conventional mappings arises early in the learning of these artificial lexicons; given sufficient training, conventional systematic lexicons become equally well learnt, suggesting that at least some previous studies may have overestimated the importance of motivatedness or iconicity for language learning, where systematic structure might be sufficient.

Furthermore, Experiment 5 shows that partially motivated lexicons exhibit the lowest rates of learnability, suggesting that the presence of both motivated and conventional associations between signals and meanings interferes with learning.

This finding runs counter to the prediction that motivated associations might bootstrap the learning of conventional ones. In order to further explore this finding, in Experiment 6 we investigate the learnability of lexicons mapping words to shapes that differ in both shape and size, allowing us to independently manipulate the motivatedness of mappings from consonants to shape (e.g. in a motivated lexicon, plosives map to jagged shapes; in a conventional lexicon, fricatives map to jagged shapes) and vowels to size (e.g. in a motivated lexicon, high vowels map to small shapes; in a conventional lexicon, mid-vowels to small shapes). In line with the results from Experiment 5, we find that lexicons that exhibit a mix of motivated and purely conventional mappings are hardest to learn, again counter to the bootstrapping hypothesis.

4.3 Experiment 5

We conducted an artificial language learning experiment using a paradigm where participants learned associations between novel words and images, where those associations were either motivated or conventional. The stimuli used in this experiment were similar to those used in Nielsen & Rendall (2012), which explored the learnability of motivated and counter-motivated lexicons. Participants were assigned to one of four conditions. The target lexicons across all four conditions were systematic, but differed in their level of motivatedness. Following the experiments reviewed above (Nielsen & Rendall, 2012; Aveyard, 2012; Monaghan et al., 2012),

participants in the Motivated condition attempted to learn a systematic lexicon that was consistent with known sound-symbolic biases, while participants in the Counter-Motivated condition attempted to learn a lexicon which was incongruent with those same biases. Participants in the Conventional condition attempted to learn a conventional systematic lexicon, using forms that were neither motivated nor counter-motivated. Finally, participants in the Partially Motivated condition attempted to learn a systematic lexicon which mixed motivated and conventional mappings.

4.3.1 Methods

Participants

Participants were 63 students and graduates recruited from the Student and Graduate Employment recruiting service at the University of Edinburgh and were assigned randomly to each of the four experimental conditions. Of the 63 participants, 39 were female and the average age of the participants was 23.65 years. All participants were proficient speakers of English and had normal hearing and normal or corrected-to-normal vision. Participants were paid £2 for their participation, which took approximately 10 minutes.

Image and Word Stimuli

The images used in this experiment were selected from two distinct categories: curved and jagged shapes, and were generated using a random shape generator (Birkbeck, 2008). The generator populates a field of a given size with a set of random initial points to determine an image seed, and then connects these points via cubic Bezier curves. Using a radially constrained methodology, these randomly generated points are joined using either straight lines, or via the migration of interpolated points to create curved versions of images using the same seed (see also Nielsen & Rendall, 2011, 2012, 2013 for similar image generation techniques). All of the images used in the study were simple black line figures presented on a white background as bitmap files with 480x480 resolutions (see Figure 4.02). The pair of images presented on each trial was created for the same image seed, and were thus maximally similar to one another despite one being curved and the other jagged.

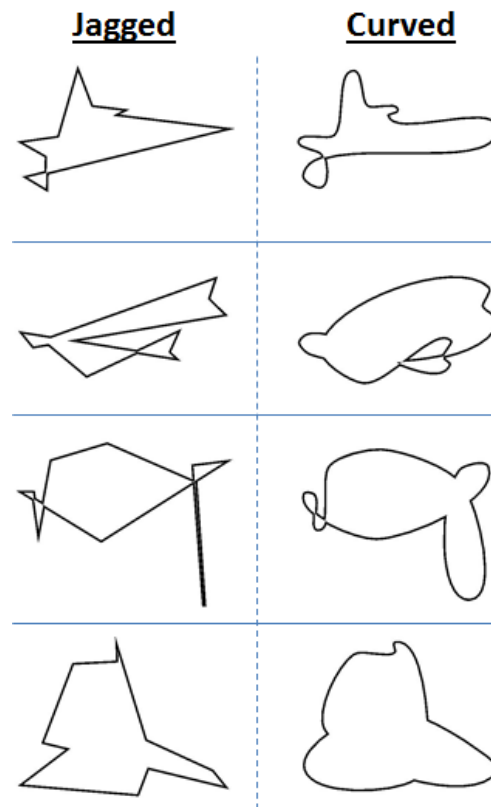


Figure 4.02- *An example of the figures used in Experiment 5. Each row represents a pair of images generated from the same initial seed.*

The words used to label these objects were all disyllabic cVcV words. For participants in the Motivated condition, labels were constructed that were congruent with previously observed sound symbolic biases: labels for curved images contained the phonemes /m/ and /n/, while labels for jagged images contained the phonemes /p/ and /t/ (see Table 4.01 for the assignment of consonants to conditions, and see below for an explanation of how these consonants were combined with vowels to form

words). In the Counter-Motivated condition, words were created identically to the Motivated condition, but assigned in the opposite (incongruent) manner, such that plosive words were assigned to curved objects and sonorant words to jagged objects. In the Conventional condition, labels were similarly systematic, but the pairings of phonemes to shapes was not motivated by previously established sound symbolic biases: two sets of phonemes were chosen (/θ/ and /ð/ vs. /ʃ/ and /ʒ/)³, and for each participant one phoneme set was paired with curved labels and the other with jagged labels.

Finally, in the Partially Motivated condition, words were created by mixing motivated and conventional phonemes: for each participant, one of the phonemes associated with each word category would be motivated (e.g. /n/ for a label for a curved object) while the other would be conventional/arbitrary (e.g. any of /θ/, /ð/, /ʃ/, or /ʒ/). These partially motivated lexicons were generated such that no single phonetic feature that varied systematically between the two label types (e.g. there was no voicing contrast that split the categories). One quarter of the labels for any lexicon in the Partially Motivated condition therefore featured only motivated

³ See the Discussion for some remarks on issues relating to differences between conditions in within-category phoneme similarity.

phonemes, one quarter featured only non-motivated phonemes, and half featured one motivated and one non-motivated phoneme.

	Motivated Condition	Conventional Condition	Partially Motivated Condition	Counter-Motivated Condition
Curved Images	m, n	θ, ð	m, θ	p, t
Jagged Images	p, t	ʃ, ʒ	p, ʃ	m, n

Table 4.01- Consonant phonemes used in Experiment 5

The vowels /ʌ/ and /ɛ/ were used across all four conditions, yielding 4 possible syllables for each label type in each condition, which were then concatenated to produce 16 possible disyllabic labels for each type of stimuli in each condition of the experiment (see Table 4.02 for examples).

	Motivated Condition	Counter-Motivated Condition	Conventional Condition	Partially Motivated Condition
Curved Images	mʌnʌ	pʌtɛ	ʃɛʒʌ	mɛmʌ
	mɛmɛ	pɛpɛ	ʒʌʃʌ	ʃʌʃʌ
	nɛmʌ	tʌtɛ	ʃʌʃɛ	mʌʃɛ
	nʌnɛ	tɛpʌ	ʒɛʒɛ	ʃɛmɛ
Jagged Images	pʌtɛ	mʌnʌ	θʌðɛ	θʌtɛ
	pɛpɛ	mɛmɛ	θɛðɛ	θɛtɛ
	tʌtɛ	nɛmʌ	ðɛðʌ	tɛtʌ
	tɛpʌ	nʌnɛ	ðɛθʌ	tɛθʌ

Table 4.02- Sample words used in Experiment 5

The phoneme segments required for the experiment were recorded by a trained male phonetician in a single continuous track and then extracted as sound files. These files were then concatenated using the SoX command line sound processing utility (<http://sox.sourceforge.net/>) to produce all possible labels for each experimental condition. Assembling the word stimuli in this way ensured an accurate and consistent presentation of the phonemes in question and also allowed for the construction of words that did not contain any stress information and where there was no influence of coarticulation. Thus, although the stimuli were still somewhat artificial, they were markedly less artificial sounding than the stimuli from Experiments 1-4.

Procedure

The experiment was conducted using an interface created with Livecode (Version 5.50, RunRev, 2012). Participants were randomly assigned to one of the four conditions. On each of 96 experimental trials, participants were presented with a pair of images, one jagged and one curved (with location on screen, left vs. right, randomized) and, after 500 ms, played one of the word stimuli via headphones. One second after the first presentation of the word, it was presented again (see Fig 4.03A for an example of a typical trial).

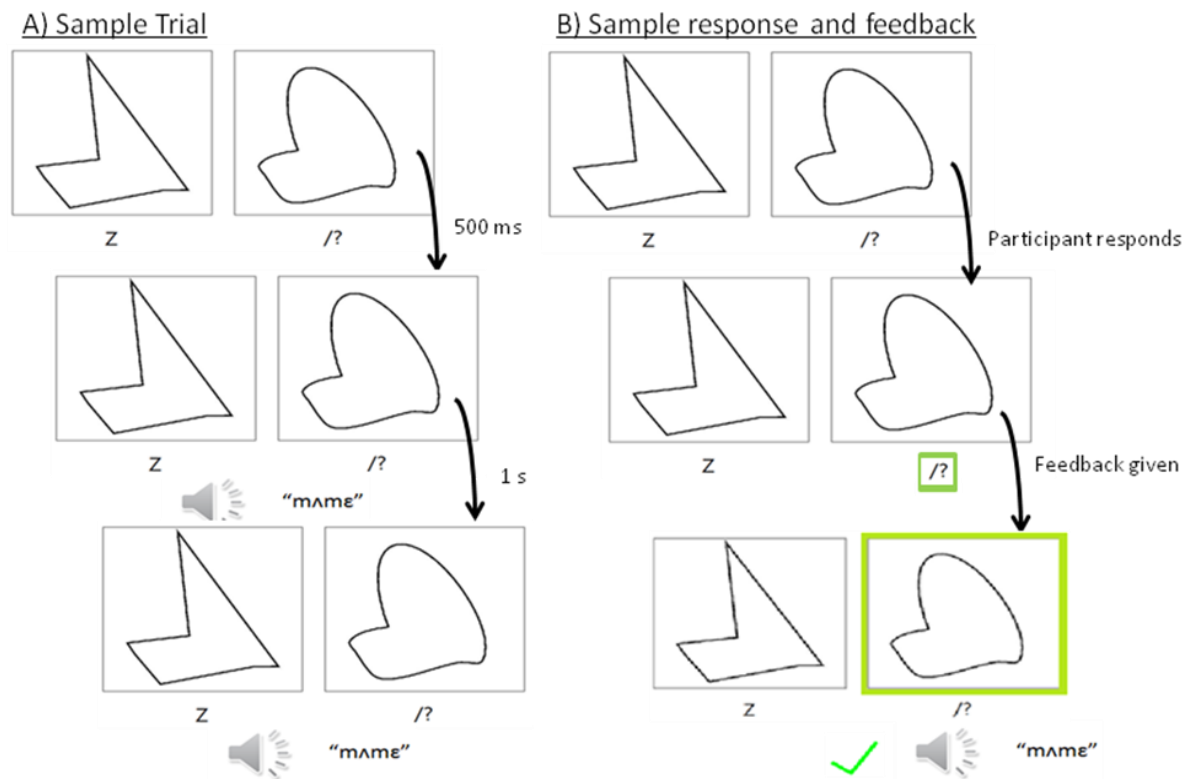


Figure 4.03- A) An example of what participants saw on each trial of the experiment.
B) Feedback given to participants after their response, showing that they responded correctly and highlighting the correct image for the presented label

On each trial participants were tasked with choosing the image that matched the label presented to them, which they did by pressing either the “Z” or “/?” key on the keyboard. Participants were provided with feedback after every trial (Fig 4.03B). If participants responded correctly, they were shown a green checkmark at the bottom of the screen, while if they responded incorrectly they were shown a red “x”. The label for the trials was played to them again, and the correct image was highlighted with a green square.

The 96 trials were split into three blocks, each of 32 trials, with each of the 32 possible labels being presented once per block, and the order of labels within blocks randomized. In each block, a given label was paired with a new pair of images, such that no image was seen more than once (i.e. there were 96 pairs of images, randomly distributed between the three experimental blocks). The lexicon that participants learned therefore provided labels for categories of images, rather than individual images: on each of the three occurrences of a label, the correct answer would be of a consistent category (e.g. curved or jagged) but not identical to the previous correct answer seen for that label.

Data Analysis

Responses for each trial of the experiment were coded for correctness and then analysed using a logistic mixed effects analysis of the relationship between correctness and lexicon type. The analysis was conducted using R (R Core Team, 2012) and lme4 (Bates, Maechler, & Bolker, 2015). We used experimental block and condition (and their interaction) as fixed effects, with by-subject random intercepts and by-subject random slopes for the effect of block; condition was dummy-coded, taking the Conventional condition as our reference level; Block was a numerical predictor, with the model intercept giving the log-odds of correct responses in block 1. P-values for fixed effects and their interaction were obtained using likelihood ratio

tests of the full model compared against the model without the effect; other p-values reported below were obtained via the normal approximation.

A second analysis, identical to the first but using linear rather than logistic regression, was conducted examining the effect of lexicon type on response times (time between the start of the audio clip being played and the participant providing their response by key press).

In addition to an omnibus analysis using Experimental Block and Condition as factors, we conducted an additional planned analysis looking at only the first eight trials for each participant- an early period of time over which we could expect any differences between the conditions to be the most pronounced.

4.3.2 Results

Performance over time in all four conditions is shown in Figure 4.04. Model comparison revealed a significant effect of experimental condition ($\chi^2(6) = 22.58$, $p < .001$). Participants in the Conventional condition performed significantly better than chance even in block 1 ($\beta = 0.98$, $SE = 0.22$, $p < .001$). Participants in the Counter-Motivated condition did not perform significantly differently from participants in the Conventional condition ($\beta = 0.01$, $SE = 0.33$, $p = .98$). Participants in the Motivated condition performed significantly better ($\beta = 0.83$, $SE = 0.32$, $p = .01$), whereas participants in the Partially Motivated condition performed significantly *worse* ($\beta = -$

0.74, SE=0.34, $p=.027$); indeed, participants in the Partially Motivated condition are not significantly better than chance in block 1 (a model taking Partially Motivated as the reference level of condition has a non-significant intercept, indicating that the log-odds of a correct answer are not significantly greater than 0, i.e. the odds are not significantly greater than 1: $\beta=0.24$, SE=0.25, $p=.34$).

Model comparison also revealed a significant effect of Block ($\chi^2(4)= 69.02$, $p<.001$) and a significant interaction between Block and Condition ($\chi^2(3)= 8.00$, $p=.046$). Performance increased over blocks in the Conventional condition ($\beta=0.92$, SE=0.17), and increased at similar rates in the Motivated and Counter-Motivated conditions (as indicated by the lack of significant interaction with Block for these conditions: Motivated, $\beta=0.12$, SE=0.25, $p= .63$; Counter-Motivated, $\beta=0.19$, SE=0.25, $p= .46$). However, performance increases marginally more slowly with Block in the Partially Motivated condition ($\beta=-0.47$, SE=0.24, $p=.052$).

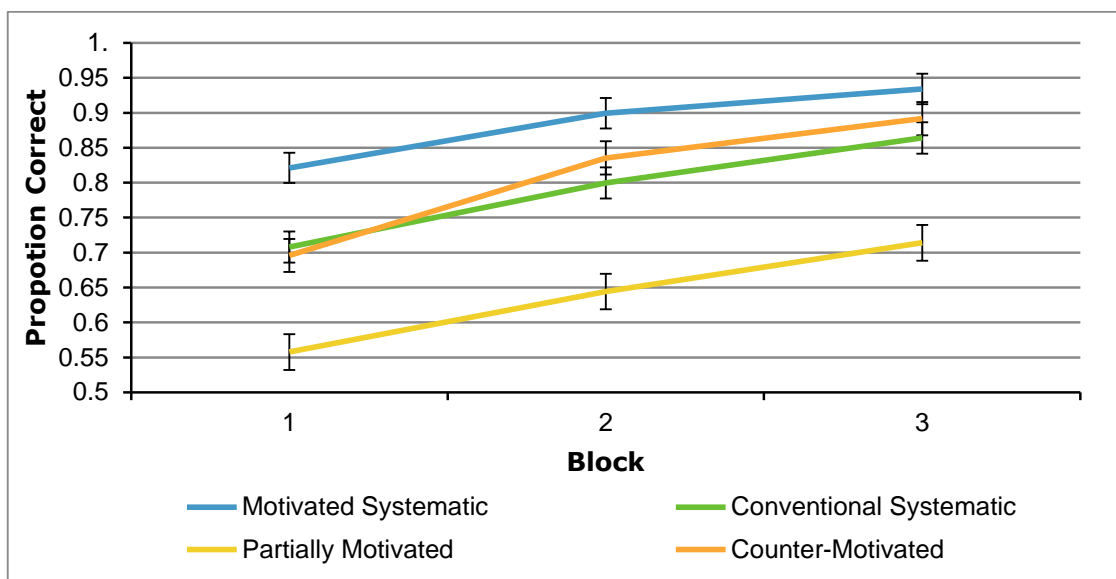


Figure 4.04 - Results of performance from Experiment 5 show a significant advantage for learners of Motivated Systematic language learners, especially in Block 1. Further, the results show partially motivated lexica to be learned least effectively, with little difference between conventional and counter-motivated lexicons. Error bars show standard error.

Our analysis of the earliest exposures, limited to the first 8 trials and with only Condition as a fixed effect, revealed a significant effect of condition ($\chi^2(3) = 11.72$, $p = .008$). Participants in the Conventional condition were not performing significantly above chance in the first 8 trials ($\beta = 0.18$, $SE = 0.20$, $p = .36$); while performance of participants in the Partially Motivated condition did not differ significantly from the Conventional condition ($\beta = -0.06$, $SE = 0.30$, $p = .83$), participants in the Motivated condition performed significantly better, and indeed substantially above chance ($\beta = 0.88$, $SE = 0.29$, $p = .003$). Participants in the Counter-Motivated condition exhibiting an intermediate level of performance: the model with the Conventional condition as the reference level indicated that participants in the Counter-Motivated condition were not significantly better than participants in

the Conventional condition ($\beta=0.42$, $SE=0.30$, $p=.15$), while a model using the Counter-Motivated condition as the reference level showed that their performance was also not significantly lower than participants in the Motivated condition ($\beta=0.46$, $SE=0.31$, $p=.14$), but was significantly above chance ($\beta=0.61$, $SE=0.22$, $p=.006$).

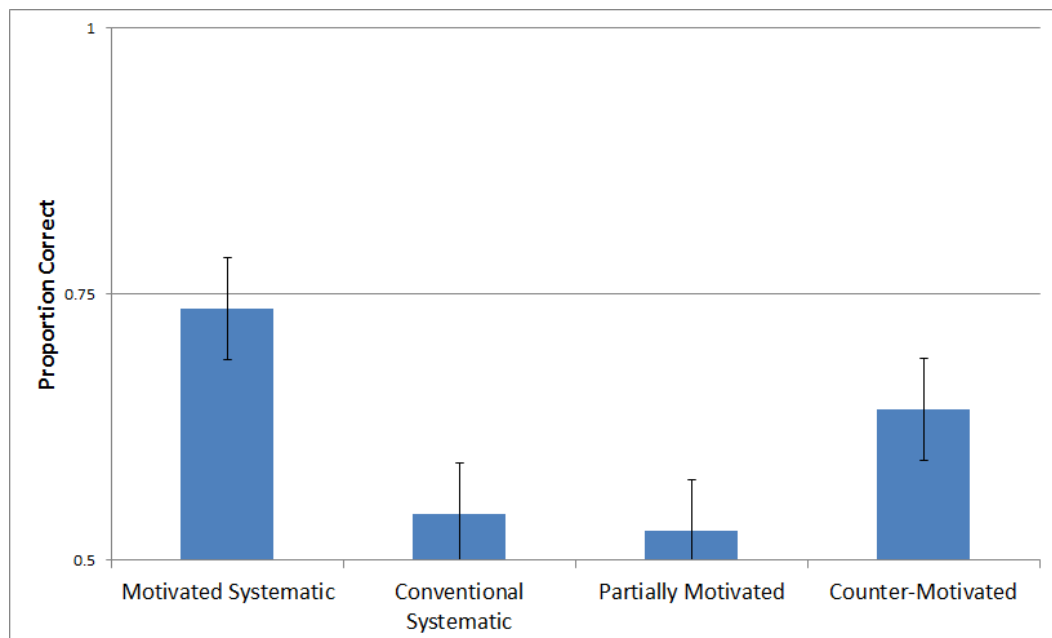


Figure 4.05- Performance in first 8 trials for Experiment 5. Motivated Systematic and Counter-Motivated language are above chance even in the first 8 trials, while systematic and partially motivated lexica are at chance. Error bars show standard error.

In our analysis of response times (see Figure 4.06), model comparison indicated a significant effect of experimental condition ($\chi^2(6)=20.99$, $p=0.002$), which was driven primarily by participants in the Counter-Motivated condition: while participants in the Conventional, Motivated and Partially Motivated provided

responses approximately equally rapidly in Block 1 (model intercept indicating response times in the Conventional condition: $\beta=2240\text{ms}$, $\text{SE}=186\text{ms}$; no significant difference in the Motivated condition, $\beta=62\text{ms}$, $\text{SE}=259\text{ms}$, $p=.31$; nor in the Partially Motivated condition, $\beta=626\text{ms}$, $\text{SE}=282\text{ms}$, $p=.24$), participants in the Counter-Motivated condition responded significantly more slowly ($\beta=930\text{ms}$, $\text{SE}=271\text{ms}$, $p<.001$).

There was a significant effect of experimental block ($\chi^2(4)= 59.83$, $p<.001$), but no interaction between condition and block ($\chi^2(3)= 1.62$, $p=.65$): response times decreased by over 300ms per block in the reference Conventional condition ($\beta=-343$, $\text{SE}=80$, $p<.001$), and similar decreases were seen in all other conditions.

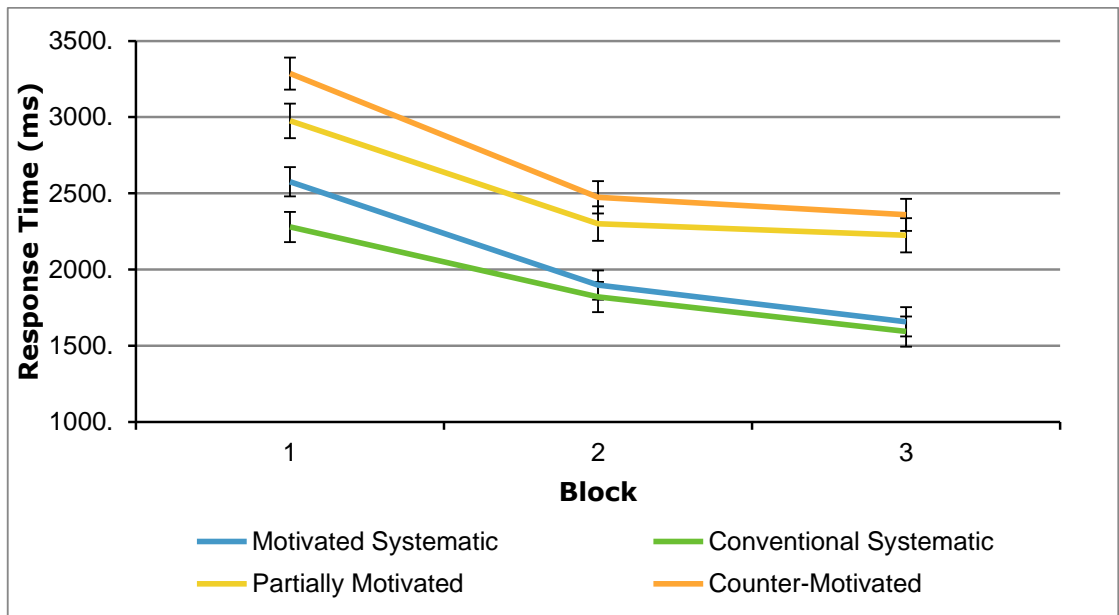


Figure 4.06- Response Time Data for Experiment 5. Despite the fact that motivated and counter-motivated languages show similar early learning advantages (in trials 1-8), and counter-motivated and systematic languages show similar patterns of learning over the course of the whole experiment, there is a significant penalty with regards to reaction time for counter-motivated lexica, suggesting that extra processing is required at some level. Error bars show standard error.

4.3.3 Discussion

Motivated connections between words and objects provide an early advantage to language learners: even with very little training, the fact that connections between words and meanings aligns with their perceptual biases allows learners to rapidly perform above chance levels, and significantly better than learners of purely conventional systematic lexicons. Learners of purely conventional systematic lexicons have no such early advantage, and thus require experience with their

artificial language; over the course of repeated exposure however, performance of participants in the conventional condition increases, with learners of both motivated and conventional lexica performing near ceiling by the third block of the experiment. This result is consistent with the literature reviewed in the introduction, highlighting the advantages of motivated mappings, and shows that this advantage persists when motivated lexicons are compared against conventional, rather than counter-motivated, lexicons.

Our other results are more surprising. First, we found, counter to our expectations based on previous experiments (Nielsen & Rendall, 2012; Aveyard, 2012; Monaghan et al., 2012) that the counter-motivated lexicon was learned as well as the purely conventional lexicon. Second, contrary to our expectations based on the bootstrapping hypothesis, we found that the partially motivated lexicon was learned worst of all.

In previous experiments that examined the difference in learnability between motivated and counter-motivated artificial lexicons, researchers have found consistently that their participants performed significantly worse at learning counter-motivated lexicons (Aveyard, 2012; Monaghan et al., 2012), and in one case, found that learners of a counter-motivated lexicon didn't even perform at rates above chance (Nielsen & Rendall, 2012). Our results are broadly consistent with this picture: over the course of the entire experiment, performance in the counter-motivated condition is lower than in the motivated condition (see e.g. Figure 4.04

and associated analyses). However, performance on the counter-motivated lexicon is generally high, and certainly no worse than performance on the conventional lexicon. This comparison between counter-motivated and purely conventional lexicons, absent in previous work, suggests that the difference seen previously between motivated and counter-motivated lexicons is likely to be driven largely by an *advantage* to motivated mappings, rather than a penalty to counter-motivated lexicons: counter-motivated lexicons seem to be no harder to learn than any other systematic lexicon.

Surprisingly, we found that counter-motivated lexicons exhibited performance intermediate between motivated and conventional lexicons in the earliest trials for each participant, and were above chance in those early trials (like participants learning a motivated lexicon, but unlike participants learning a purely conventional lexicon), suggesting that participants were able to productively use counter-motivated mappings after a very small number of exposures. One potential explanation for this finding is that the counter-motivated lexicons require only the addition of one additional step of processing, namely reversal of expectation: participants identify the ‘best’ (i.e. motivated) referent for a label, and then select the other referent. The reaction time data support this interpretation - participants in the counter-motivated condition were significantly slower than participants in the Motivated and Conventional conditions (approximately 700ms slower on average to respond on each trial). One interpretation of this data is that this extra time

required to respond reflects a conscious strategy by participants to reverse their expectations, but the deficit in response speed is equally well explained as operating subconsciously - in either case, the large response time difference is suggestive of the same kinds of explanations offered by the findings of previous studies, namely pointing to the ‘naturalness’ of motivated mappings.

Our finding that participants in the counter-motivated condition were reliably above chance, even in the first 8 trials, runs counter to the finding by Nielsen & Rendall (2012) that participants were unable to perform above chance after 12 training trials with feedback. However, Nielsen & Rendall (2012) used a signal detection paradigm, where participants responded to single images paired with labels, which may have made differences between curved and jagged images less obvious, or make explicit or implicit reversal-of-expectation approaches of the sort we see evidence for in our task less accessible for participants - the overall more modest results of Nielsen & Rendall (2012) when compared with other sound symbolism work using a 2AFC paradigm (e.g. Nielsen & Rendall, 2011; Maurer et al., 2006) also suggests this possibility. In addition to using a potentially more difficult signal-detection paradigm, Nielsen & Rendall (2012) provided a single block of training, where participants may or may not have been able to internalize appropriate rules for responding and thus simply reverted to their existing perceptual bias. In the paradigm reported here, feedback given after each trial enables participants to eventually learn to make the correct associations.

Counter to bootstrapping predictions made in the sound symbolism literature, we found that participants in the partially motivated condition performed significantly *worse* than those in the other conditions, suggesting that the concurrent use of motivated and conventional markers for a single semantic dimension (spikiness/curvedness) might be problematic for lexicon learners. This interpretation of the data runs counter to previous claims in the literature that one of the benefits of motivated associations might be that they bootstrap the learning of related conventional tokens, and thus help account not only for the learning of motivated tokens themselves, but also the rest of the lexicon.

Although this is an intriguing possibility, one potential alternative explanation for the difficulty posed by our partially-motivated lexicon is that it is driven by phonological feature similarity, rather than the mixing of motivated and conventional mappings. In our conventional lexicon, for example, the phonemes θ and δ are very similar, differing only in their voicing, and refer to the same category of referents. In one possible instantiation of the partially motivated condition however, the two phonemes θ and m (differ in voicing, sonority, nasality of stricture, and place of articulation) might be used to refer to shapes drawn from single category. Participants in this condition of the experiment are therefore faced with a more difficult task of mapping two dissimilar phonemes to a single category. However, this seems unlikely to be a full explanation. Firstly, the phonemes used in the motivated lexicons in our experiment are more distinct within-category than our

conventional phonemes (p and t are more dissimilar than Θ and δ), yet are learnt more successfully. Secondly, in other work using a similar artificial language learning paradigm (Nielsen et al., in prep), we find that adult learners experience no difficulties in learning lexicons with highly dissimilar within-category phonemes, at least in the case where both phonemes are conventionally systematic with relation to the category that they mark.

One additional possible explanation for the deficit in learnability of partially motivated lexicons is that the bootstrapping of conventional label-meaning mappings by the presence of motivated mappings can only be effective *across* meaning dimensions, rather than within them. In our partially motivated lexicon, a mix of conventional and motivated forms are used to convey a single semantic dimension, spikiness versus curvedness. It could be that this sort of competition between motivated and non-motivated mappings is problematic, and that motivated mappings on one meaning dimension might bootstrap the learnability of conventional mappings on a second, unrelated meaning dimension, or at least not interfere with learning in other dimensions. To explore this possibility, we conducted a second experiment where image stimuli varied along two dimensions (shape, as in Experiment 5, and size), both of which have known sound-symbolic associations.

4.4 Experiment 6

As in Experiment 5, we conducted an artificial language learning experiment where participants learned associations between novel words and images that were either motivated or conventional. The stimuli used in the experiment were similar to those used in Experiment 5, but featured a larger space of vowels, and images varied in size as well as in jaggedness. Dating back to at least Sapir (1929), previous work has demonstrated in both English (Johnson, 1967) and a number of other languages (Gebels, 1969; Huang, 1969; Kim, 1977; Malmberg, 1964) that high front vowels are associated with small size and low back vowels are associated with large size, making size a suitable second dimension which can be encoded linguistically in a motivated or purely conventional manner. We trained and tested participants on four lexicons, in a between-subjects 2x2 design where we independently manipulate whether shape and size are linguistically encoded in a motivated or purely conventional manner; this design therefore allowed me to explore whether partially-motivated lexicons are always harder to learn, or whether this disadvantage only relates to competition between motivated and conventional mappings on a single semantic dimension.

4.4.1 Methods

Participants

Participants were 48 students and members of the public recruited from the SAGE recruiting service at the University of Edinburgh, and were assigned randomly to each of the four experimental conditions. Of the 48 participants, 32 were female and the average age of the participants was 22 years. All participants were proficient speakers of English and had normal hearing and normal or corrected-to-normal vision. Participants were paid £2 for their participation, which took approximately 20 minutes.

Image and Word Stimuli

The images used in this experiment were created using the same procedure outlined in Experiment 5, but in addition to varying in jaggedness they also varied in size.

Small images were presented in 240x240 resolution, whereas large images were presented in 480x480 resolution, thus small images were $\frac{1}{4}$ of the area of their larger counterparts (see Figure 4.07).

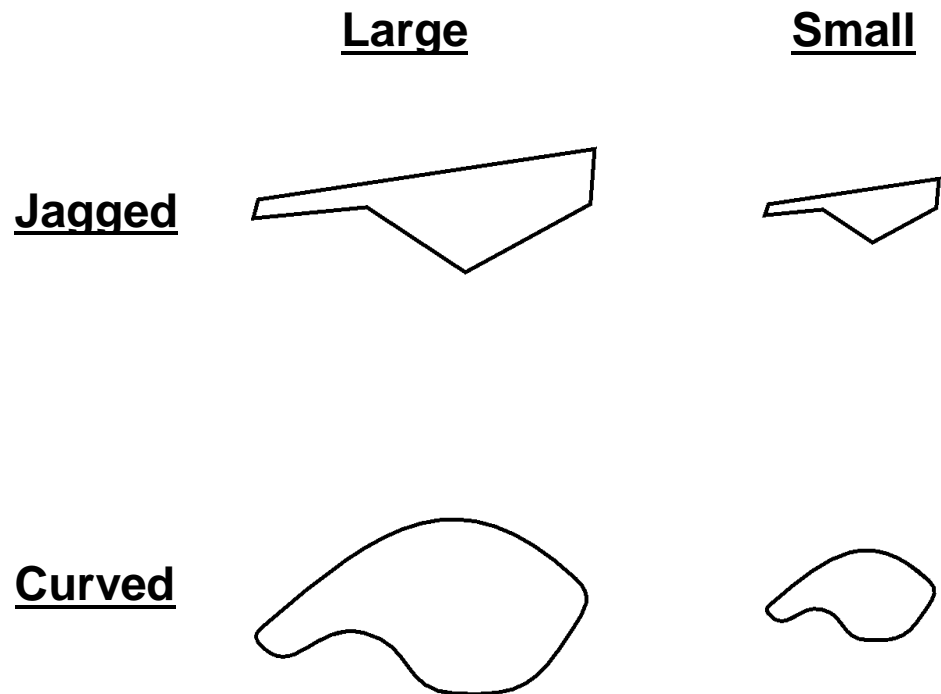


Figure 4.07- Examples of images used in Experiment 6, where images vary on both jaggedness and size.

The words created for Experiment 6 were all disyllabic cVcV words, as in Experiment 5. As in Experiment 5, motivated labels were constructed using phonemes congruent with known sound-symbolic biases. Labels for curved images contained the phonemes /m/ and /n/, while labels for jagged images contained the phonemes /p/ and /t/. Vowels were chosen for the motivated sound-symbolic mapping based on previously observed biases where large objects are typically associated with low back vowels and smaller objects with high front vowels (Sapir,

1929; Huang, 1969). Thus, we chose the vowels /a/ and /v/ for large images and the vowels /i/ and /y/ for small images.

For non-motivated, conventional conditions of the experiment, we selected labels featuring phonemes whose association to shapes was not motivated by any previously established sound symbolic biases. Thus, for participants for whom the sound-shape mapping was non-motivated, two sets of consonants were chosen (/θ/ and /ð/ vs. /ʃ/ and /ʒ/) and for each participant one set was paired with curved labels and the other with jagged labels; for participants for whom the sound-size mapping was non-motivated, two sets of vowels were chosen (/I/ and /ʌ/ vs. /ʊ/ and /ɛ/) and for each participant one set of these vowels was paired with small images and the other with large images (see Table 4.03).

		SHAPE MOTIVATED		SHAPE CONVENTIONAL		
SIZE MOTIVATED		Large	Small		Large	Small
	Jagged	pato	pyti	Jagged	Θaðv	Θyði
	Curved	mano	myti	Curved	ʃaʒv	ʃyʒi
SIZE CONVENTIONAL		Large	Small		Large	Small
	Jagged	pItʌ	pʊts	Jagged	ΘIðʌ	Θʊðɛ
	Curved	mInʌ	mʊts	Curved	ʃIʒʌ	ʃʊʒɛ

Table 4.03- A sample of labels used in Experiment 6.

This set of images and sounds allowed me to independently manipulate whether shape or size were mapped to sound in a motivated or purely conventional manner,

yielding a 2x2 between-subjects design. Participants in the Shape Motivated – Size Motivated condition (or Motivated-Motivated for short) were trained on labels which were congruent with sound-symbolic associations for both consonant-shape and vowel-size. Participants in the Shape Conventional-Size Conventional (Conventional-Conventional) condition were taught associations between signals and meanings that were conventionally systematic for both vowel-size and consonant-shape. In the remaining two conditions, one of the shape features was coded sound symbolically, and the other conventionally (Shape Motivated-Size Conventional; Shape Conventional-Size Motivated).

In each of these four experimental conditions, the two vowels and two consonants for each type of image were combined to create 4 possible syllables, which could then be concatenated in all combinations to produce 16 possible labels for each type of stimuli in each condition of the experiment. In Shape Conventional conditions, the number of participants for which the phonemes /ʃ/ and /ʒ/ were used to label curved objects was counterbalanced across participants; similarly, in Size Conventional conditions the number of participants for which the phonemes /I/ and /ʌ/ were used to label small objects was counterbalanced across participants. The phoneme segments required for the experiment were recorded by a trained phonetician in a single continuous track and then extracted as sound files. These files were then concatenated using the SoX command line sound processing utility to produce all possible labels for each experimental condition.

Procedure

The experimental procedure was closely matched to Experiment 5. On each of 192 experimental trials participants were presented with a pair of images, and, after short delay, played one of the word stimuli via headphones. One second after the first presentation of the word, it was presented again.

On each trial participants were tasked with choosing the image that matched the label presented to them, which they did by pressing either the “Z” or “/?” key on the keyboard, and were provided with feedback after every trial.

There were three types of trials in Experiment 6. Both Different trials presented pairs of images that were different on both size and shape: thus, participants would be able to answer correctly on Both Different trials if they had learned either type of association (vowel-size or consonant-shape). Size Different trials presented pairs of images that varied only on size (and had the same shape); thus, to answer correctly on these trials participants needed to be familiar with the vowel-size mapping in their lexicon. Finally, Shape Different trials presented pairs of images that were identical in size, but differed in shape (one image was curved and the second jagged); to answer correctly on these trials participants needed to be familiar with the consonant-shape mapping in their lexicon (Figure 4.08).

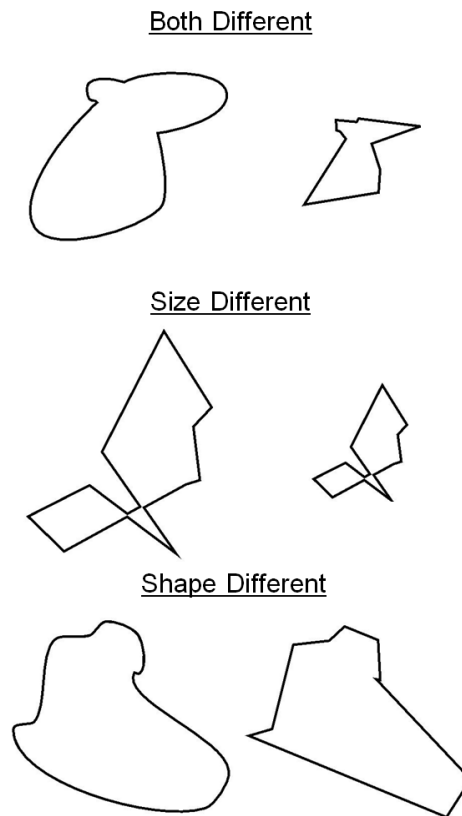


Figure 4.08- *The three trial types used in Experiment 6.*

In each of three blocks of trials there were 64 total trials: 24 Both Different trials, 20 Size Different trials, and 20 Shape Different trials. On each trial which image was presented on each side was randomized. The experiment was conducted using an interface created with Livecode 5.50.

Data Analysis

As in Experiment 5, responses for each trial of the experiment were coded for correctness and then analysed using a logistic mixed effects analysis of the relationship between correctness and lexicon type. The analysis was conducted using R (R Core Team, 2012) and lme4 (Bates et al., 2015). Omnibus analyses indicated several interactions involving trial type and block, and in the interests of clarity we therefore analyze all three trial types separately. For each trial type we use experimental block, consonant mapping (conventional versus motivated association with shape), vowel mapping (conventional versus motivated association with size), and their interactions, as fixed effects, with by-subject random intercepts and by-subject random slopes for the effect of block, consonant mapping and vowel mapping. Block was a numerical predictor, with the model intercept giving the log-odds of correct responses in block 1; consonant mapping and vowel mapping were dummy-coded with Conventional as the reference level, yielding models whose intercepts indicate performance on lexicons with conventional consonants and conventional vowels (at block 1). P-values for fixed effects and their interaction were obtained using likelihood ratio tests of the full model compared against the model without the effect; other p-values reported below were obtained via the normal approximation.

4.4.2 Results

Performance over time in all four conditions for each of the three trial types is shown in Figures 4.09-4.11.

Both Different trials

On Both Different trials, participants in the Conventional-Conventional condition performed above chance even in block 1 (as indicated by a significant model intercept: $\beta=1.52$, $SE=0.26$, $p<.001$; note that experimental blocks are longer in Experiment 6 than Experiment 5). Model comparison indicated a significant effect of block ($\chi^2(4)=37.24$, $p<.001$), with performance increasing markedly on Both Different trials as participants progressed through the experiment ($\beta=0.59$, $SE=0.19$); there was also a significant effect of consonant mapping ($\chi^2(4)=10.11$, $p=.039$) and a marginal interaction between consonant mapping and block ($\chi^2(2)=4.8295$, $p=0.089$): inspection of the estimates of slope provided by the full model suggests that these effects are driven by the fact that performance on lexicons with motivated consonants increased more rapidly with block (as indicated by the interaction between consonants and block: $\beta=0.55$, $SE=0.30$).

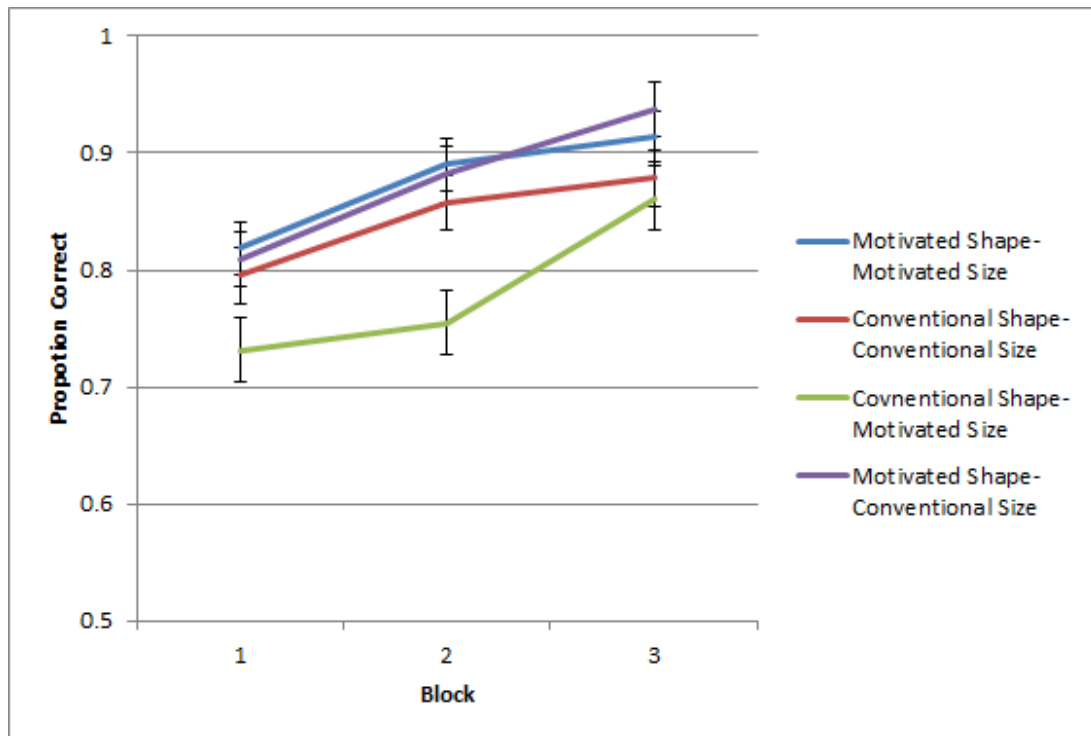


Figure 4.09- Performance on trials where both shape and size are relevant features. There is a steady increase with block in all conditions, and performance on lexicons with motivated mappings for shape increase more rapidly. Error bars show standard error.

Shape Different trials

On Shape Different trials, participants performed significantly better than chance even in block 1 (as indicated by a significant model intercept: $\beta=1.31$, $SE=0.32$, $p<.001$). Model comparison indicated a significant effect of block ($\chi^2(4)= 34.33$, $p<.001$), with performance increasing as participants progressed through the experiment at a rate similar to that seen in Both Different trials ($\beta=0.86$, $SE=0.20$). Model comparison also suggest an interaction between vowel mapping and block

($\chi^2(2)=5.5719$, $p=.06$): performance on lexicons with motivated vowels increased less rapidly ($\beta=-0.62$, $SE=0.24$) than the reference level. Note that vowels are not relevant to performance on Shape Different trials, suggesting that the presence of motivated size-vowel mappings in the lexicon interferes with learning of conventional shape-consonant mappings.

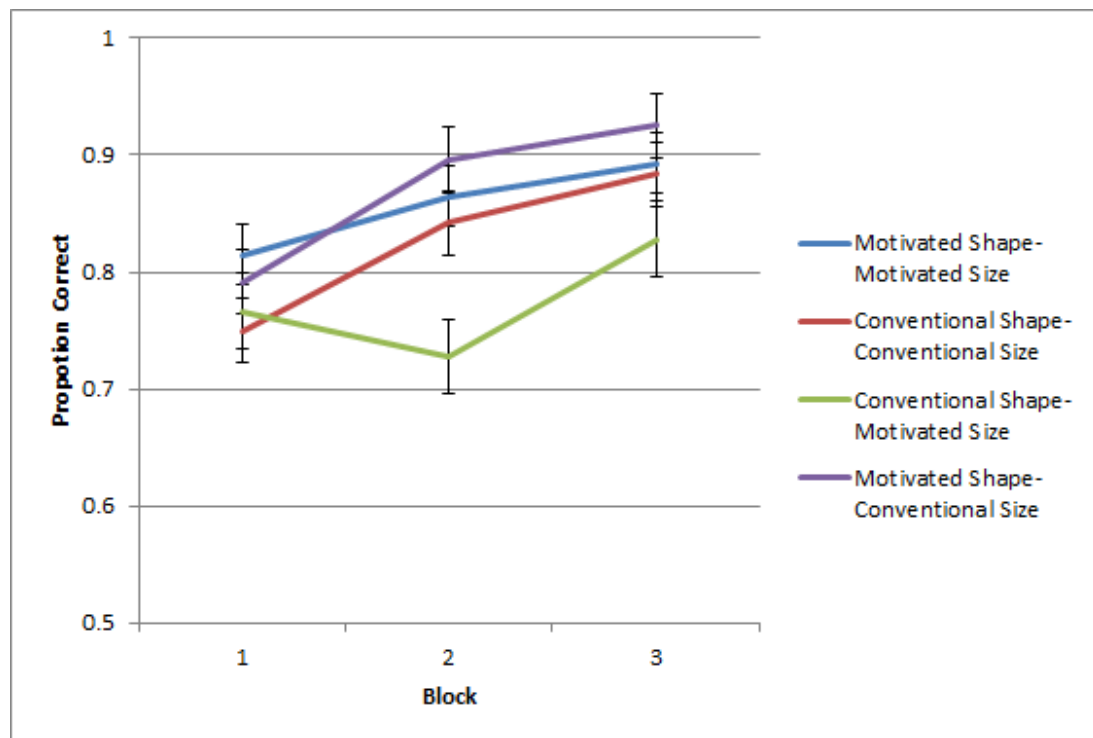


Figure 4.10- Performance on Shape Different trials, where shape is the only relevant feature for responding. The presence of motivated size-vowel mappings interferes with learning of conventional shape-consonant mappings, leading to marginally slower increase in performance over blocks. Error bars show standard error.

Size Different trials

Participants in the Conventional-Conventional reference level did not perform significantly better than chance in block 1 (as indicated by a non-significant model intercept: $\beta=0.02$, $SE=0.16$, $p=.15$). Model comparison indicated a significant effect of block ($\chi^2(4)=34.33$, $p<.001$), with performance increasing as participants progressed through the experiment ($\beta=0.46$, $SE=0.11$), albeit at a slower rate than seen in the other trial types. Model comparison also indicated that the inclusion of all other fixed effects and their interactions significantly improved model fit (consonants: $\chi^2(4)=9.74$, $p=.045$; consonants x block: $\chi^2(2)=9.27$, $p=.010$; consonants x vowels: $\chi^2(2)=6.34$, $p=.042$; vowels: $\chi^2(4)=13.08$, $p=.011$; vowels x block: $\chi^2(2)=6.49$, $p=.039$; consonants x vowel x block: $\chi^2(1)=6.23$, $p=.013$). Inspection of the full model suggests that this is driven by two effects. Firstly, in the mirror-image of the interaction seen in Shape Different trials, participants learning lexicons with motivated consonants but conventional vowels failed to increase in performance over Blocks (as indicated by the negative slope for the interaction between consonants and block, $\beta=-0.43$, $SE=0.15$; note that the magnitude of this interaction is comparable with the effect of block for the reference level, indicating that these participants did not improve with block). Again, since consonants are irrelevant to performance on Size Different trials, this effect can only be explained as a consequence of interference between (irrelevant) motivated mappings for shape

and the learning of a conventional mapping for size. Secondly, the full model exhibits a three-way interaction between consonant mapping, vowel mapping and block ($\beta=0.77$, $SE=0.28$), indicating that this interference effect is specific to learning conventional vowel-size mappings in the presence of motivated consonant-shape mappings; if *both* vowels and consonants are motivated, learning proceeds at a rate equivalent to or in excess of that seen for purely conventional lexicons.

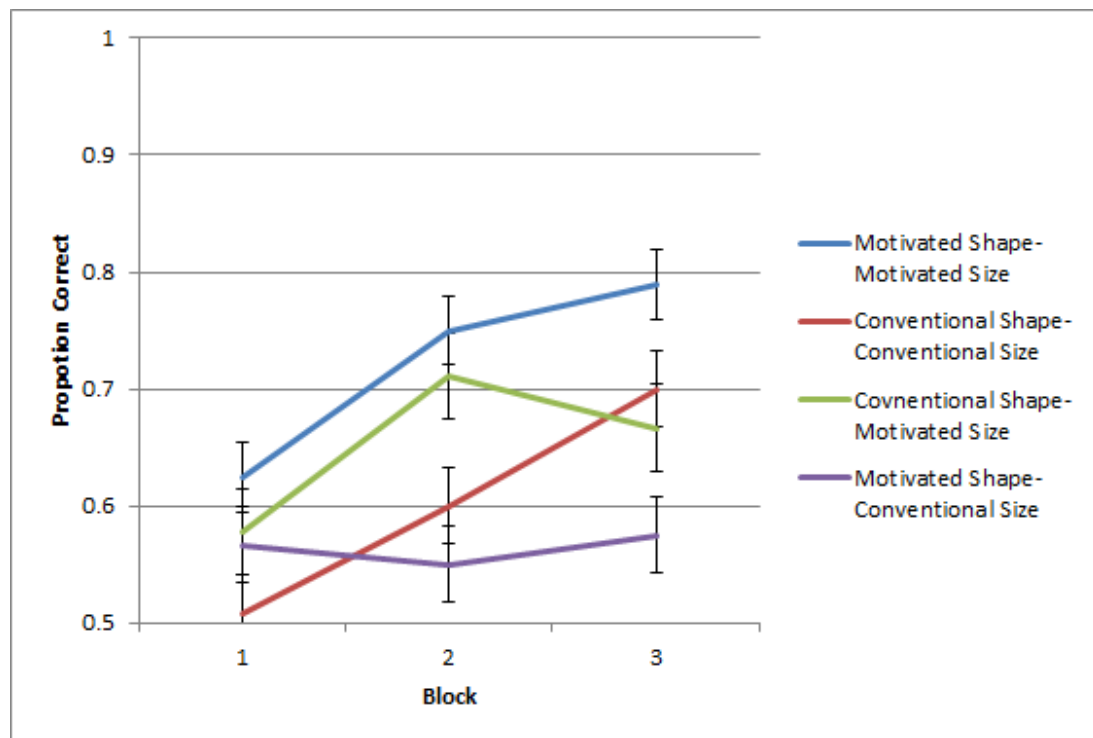


Figure 4.11- Performance on Size Different trials, where size is the only relevant feature for responding. Overall performance on these trials is substantially lower than in Both Different and Shape Different trials. The data also suggest a similar interference effect to that seen on Shape-Different trials, with the presence of irrelevant consonant-shape mappings interfering with the learning of conventional vowel-size associations, as seen in the lack of improvement over blocks in the Motivated Shape – Conventional Size lexicon. Error bars show standard error.

4.4.3 Discussion

Our results are consistent with the extensive literature on the primacy of shape in word learning (e.g. Landau, Smith & Jones, 1988; Samuelson & Smith, 2005): performance on Size Different trials, which require learning a system mapping words to object size, is markedly lower than on Shape Different trials (which involve learning mappings between words and shape), and performance on Both Different trials, where participants can exploit either word-shape or word-size mappings, pattern with Shape Different trials, suggesting that participants preferentially use word-shape mappings in this task.

In common with Experiment 5, we see some advantages for motivated lexicons relative to the baseline provided by purely conventional lexicons, although given the increased block length in Experiment 6, these advantages are less marked – performance on Both Different trials improves more rapidly when the mapping from consonants to shape is motivated. However, the clearest result from Experiment 2, consistent with the results from Experiment 5, is that there is a learning penalty associated with languages which exhibit a mix of motivated and conventional mappings, even when these mappings apply to orthogonal semantic dimensions. Subtly, however, this effect is seen through the influence of *irrelevant* motivated mappings in the lexicon, and impacts primarily on the learning of conventional mappings: on Shape Different, irrelevant motivated coding of size

results in reduced learning of conventional coding of shape; on Size Different trials, irrelevant motivated coding of shape results in reduced learning of conventional coding of size, and furthermore (again reflecting the primacy of shape over size in word learning), motivated mappings for size only facilitate learning if shape is also (irrelevantly) coded in a motivated fashion. These results, taken together with those of Experiment 5, suggest that the most learnable lexicons should be *consistent*: either consistently motivated, or consistently conventional. Mixing of motivated and conventional mappings impairs learning, a potentially problematic finding for the bootstrapping hypothesis.

4.5 General Discussion and Conclusions

The work presented here generally supports previous work examining the learnability advantages of motivated associations, but highlights that the typical comparison to counter-motivated lexica is not a fair one. If, for example, one wanted to make the claim in English that words in the gl- phonaestheme cluster were motivated, and thus easier to learn, a proper comparison would need to account for the fact that the word cluster is also systematic. Thus, some of the reported learning benefits that arise from supposedly motivated associations between words and meanings might actually be explained either jointly or entirely by the fact that those word-meaning mappings are also systematic.

Although in Experiment 5 learners of motivated systematic languages outperform those who learn conventional systematic languages, the main benefit for motivated connections between words and meanings seems to be primarily due to the fact that taking advantage of those associations requires no learning – even in the very earliest trials participants are able to respond correctly, suggesting that their perceptual bias allows for the productive use of naïve intuitions that can be especially beneficial when encountering new words.

One finding of previous research (Aveyard, 2012; Monaghan et al., 2012; Nielsen & Rendall, 2012) suggests that counter-motivated languages are more difficult for participants to learn, but in Experiment 5 we found that there was no real penalty for learners of counter-motivated lexica – they were able to learn the rules of their language nearly as well as those learning motivated languages and no worse than learners of conventional languages. This suggests that although counter-motivated associations go against perceptual biases, they are nonetheless still systematic and thus can be learned relatively easily. Intriguingly, however, participants who learned a counter-motivated language were much slower to respond, even though they were overall equally accurate. This suggests an extra level of processing to invert naïve word-meaning expectations, although whether this process is a conscious or subconscious one is currently unknown (however, the application of techniques from recent fMRI studies, e.g. Kanero et al., 2014, might be able to shed light on this question).

One of the most surprising findings to come out of Experiment 5 was that partially-motivated languages were hardest to learn - performance on these lexicons was worse than either the motivated or conventional systems from which they were created. This result suggests that an optimal system might need to be consistent with respect to its use of motivated connections between words and meanings. The results of Experiment 6 provide further support for this possibility, where the presence of irrelevant motivated associations actually impairs the ability of participants to productively apply rules for conventional associations between phonemes and meanings on a relevant dimension.

Recent findings have suggested that the learning benefits of systematic associations between words and meanings are leveraged early in language acquisition (Monaghan et al., 2014), which is especially interesting given the findings of other research which suggest that systematic lexicons are easier to learn when the size of the lexicon is small (Gasser, 2004; Monaghan et al., 2011; Nielsen et al., in prep). Monaghan et al. (2014)'s findings with regards to the systematicity of early acquired words is agnostic as to whether such associations are conventional or motivated, but the fact that languages take advantage of structure where it is most beneficial for learning is promising for proponents of bootstrapping hypotheses. As a culturally transmitted system that persists through a repeated cycle of learning and use, we expect that languages will evolve to become increasingly learnable and/or increasingly communicatively functional (Kirby, Cornish, & Smith, 2008; Silvey,

Kirby, & Smith, 2015; Winters, Kirby, & Smith, 2015; Kirby, Tamariz, Cornish & Smith, 2015); thus, if *motivatedness*, *systematicity*, or both allow for increased learnability we should expect them to be incorporated into the lexicon by the process of language transmission.

Although evidence showing that early acquired words are more systematic than their later acquired counterparts is promising for bootstrapping hypotheses, our results regarding partially motivated languages run counter to the these hypotheses. Overall, the results from both experiments suggest that optimal lexica should be consistent with regards to their use of motivated word-meaning associations, with inconsistent application of these types of associative rules impairing learnability overall. Given suggestions regarding the confusability of systematic lexica as they become more saturated (i.e. Gasser, 2004) the possibility that mixed lexica are problematic might account for a number of findings in natural languages. Specifically, if lexica need to be sufficiently large that they will inevitably become confusable, it may be the case that an early-established motivated core to a category becomes problematic and its impact on the lexicon overall needs to be mitigated.

One criticism that has been leveled against the Bouba-Kiki effect and other examples of connections between words and meanings that are motivated by the perceptuo-cognitive organisation of language speakers (some of which might be shared with other species and align with communicative pressures more generally,

cf. Owren & Rendall, 2001) is that there is little evidence that these biases are manifest in natural languages like English. Our findings regarding mixed-motivated languages might speak to this directly – because naïve expectations about sound-meaning correspondences do not operate productively when they are not applied uniformly, one might expect their impact on the lexicon as a whole to be somehow minimized. One possible mechanism for minimizing the negative impact of sound-symbolic tokens on learning of other tokens is to somehow isolate motivated tokens as a special case. In languages like Japanese, this might explain the presence of mimetic or ideophonic expressions, which are more frequent in child-directed speech and which have their own phonological and syntactic properties (Imai et al., 2008). In languages that do not make use of mimetics and ideophones, like English, the same insulation of sound-symbolic cues from the rest of the lexicon might occur at the level of speech prosody, rather than through the use of perceptual biases associated with phonemes themselves, accounting for the use of exaggerated pitch contours and durations in child directed speech.

On the other hand, one possible interpretation of the results presented here is that conventional systematic associations between words and meanings are ultimately just as effective as motivated ones, and thus we might not expect motivated mappings to crop up often in natural languages – although the presence of irrelevant sound-symbolic associations negatively impacts the learning of conventional systematic associations, there is no evidence that conventional

systematic associations induce the same kinds of learning penalties – thus, they might be able to exist in natural lexica without the need to be insulated from the rest of the lexicon by the kinds of processes outlined above. This is certainly possible, and might explain the presence of presumably conventional clusters like the *gl-* phonaestheme in English.

There are at least two ways in which the presence of sound-symbolic tokens might bootstrap the acquisition of language more generally, and future research should be mindful to make specific predictions about how such bootstrapping might work. The simplest version of a bootstrapping hypothesis that we put forward here (i.e. that the presence of motivated tokens frees up memory/effort to learn other associations) is not supported by the results of the two experiments presented here, and other work (Nielsen et al., in prep) similarly suggests that this version of bootstrapping does not align with data from artificial language learning experiments. The bootstrapping hypothesis put forward by Imai & Kita (2014) relies on the idea that the presence of sound-symbolic tokens provides a referential bootstrap that allows young language learners to establish reference. Referential bootstrapping seems like a more promising explanation for the potential benefits of motivated word-meaning associations, but given that adult experimental participants have already established concepts and categories it is difficult to test the degree to which this type of bootstrapping might influence learning. Future work should explore these bootstrapping mechanisms in children or by creating

experimental manipulations that can explore the impact of sound symbolism on the learning of novel categories (Thompson et al., 2014).

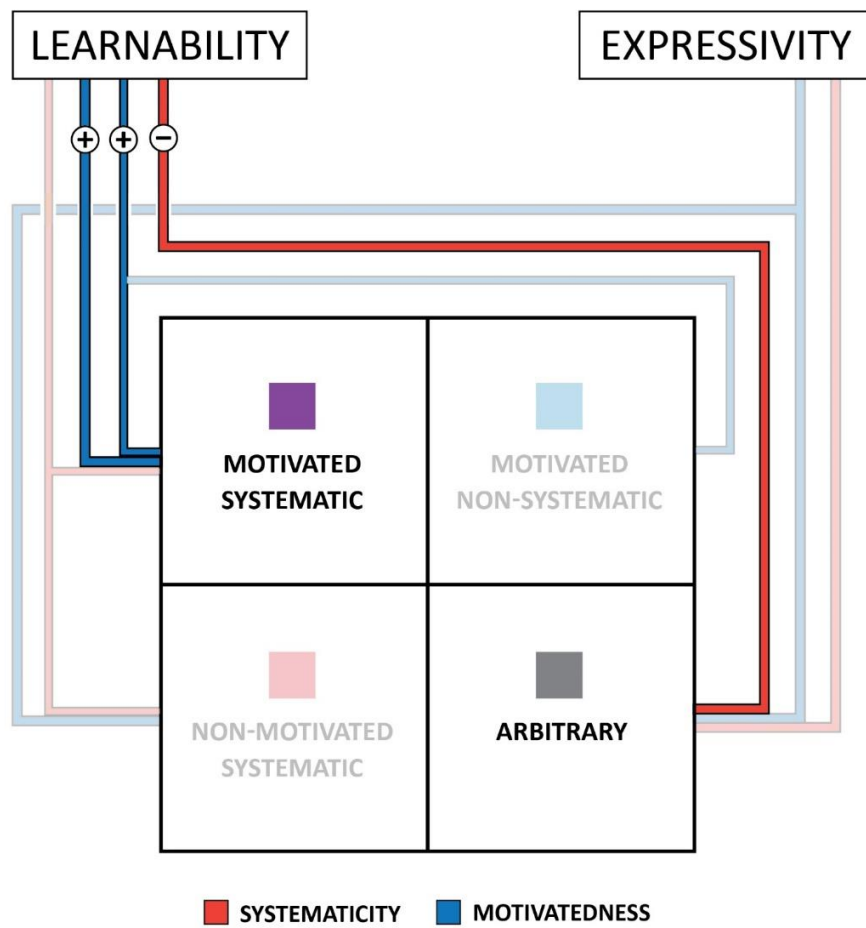
4.5.1 Conclusions

First, in line with a number of previous findings, we find an early learnability advantage for motivated lexicons. Second, although conventional and counter-motivated lexicons do not benefit from this early boost, they are subsequently learned at the same rate as motivated lexicons, and indeed exhibit very similar levels of performance in our task, suggesting that systematic counter-motivated mappings are no harder to learn than purely conventional ones. Finally, in both experiments we found a novel effect where the presence of sound-symbolic mappings interferes with the learning of conventional associations, which we speculate might be connected to a number of features of natural languages.

Afterword

The results from Experiments 5 and 6, presented in this chapter, provide a number of crucial pieces for the central argument of this dissertation, as well as pointing towards a number of areas for future research and potentially helping explain the distribution of motivated sound-meaning associations in natural languages. We found that systematic languages of all types allowed for successful categorisation, but that motivated associations between words and meanings made this systematic mapping apparent from the very earliest trials without any learning. This finding allows us to take advantage of motivated systematic associations in Chapter 5, where I will explore more directly than in previous chapters the effect that decreasing contrastiveness has on learning.

Chapter 5- Growing Lexicon Experiment



5.1 Background and Rationale

As discussed in the introduction and elaborated upon further in Chapters 2 and 3 the central question of the effect that systematicity and motivatedness have on language learning seems to hinge on the balance between the benefits accrued to those non-arbitrary mappings between words and meanings and the penalties to learnability that they might induce as similar labels become increasingly confusable. To rehearse, systematic associations between sets of words and sets of meanings allow language learners to make generalizations that positively influence the learnability of those associations- the fact that there are reliable cues to meaning in the structure of words allows for those word-meaning pairs to be learned more easily. The results of Chapter 4 further suggest that this is the case regardless of whether such systematic associations between sets of words and sets of meanings are motivated or not, with an additional small benefit to motivated mappings in the earliest stages of learning. In addition to the benefits of systematicity however, the similarity of labels to one another can have a negative impact on learnability: similar words are more readily confused, especially when they are mapped to similar meanings, and thus individuation learning can be more difficult for systematic language learners than for those learning arbitrary word-meaning mappings.

Despite experimental findings suggesting that systematicity might, in some situations, have a negative impact on learnability, the suggestion that motivated

and/or systematic associations between words and meanings might bootstrap the acquisition of the arbitrary majority of the lexicon is raised commonly in the literature surrounding these research areas. Most often, the method by which the acquisition of motivated word-meaning mappings might bootstrap the acquisition of later arbitrary tokens is left unstated (but, cf. Imai & Kita, 2014) - the fact that there is evidence that early-acquired words are more motivated than later-learned arbitrary words (Monaghan et al., 2014), and that both children and adults benefit from motivatedness when acquiring those new words (Nygaard et al., 2009, Imai et al., 2008) are offered up as evidence for the bootstrapping of arbitrary associations by motivated ones, but to date there has been no direct test of any version of a sound-symbolic bootstrapping hypothesis. This chapter presents an experiment designed to directly test this hypothesis.

5.1.1 Signal Space Saturation

To discuss contrastiveness and signal space saturation in the context of previous experiments and the experiment presented here, it is best to first rehearse what the terms mean. The idea of signal space saturation is one that is not easy to quantify for natural languages, where the overall dimensionality of the language system is unknown (see chapters 1 and 2), but in the experiments presented in this dissertation is much easier to conceptualise. If, for example, we create rules for assembling words such that a small subset of consonants and vowels are used, and words are all of a

certain length, we arrive at an absolute size for the number of possible words in that artificial language. So, for a language where there are 4 possible consonants and 4 possible vowels, with each word being a cVcV disyllable, we have 4^4 (256) possible words. In this case, the signal space saturation of a language chosen from those possible words is simply a function of how many words are chosen: we might have a relatively unsaturated signal space, where we choose to use, for example, 8 labels from the possible space, or a much more saturated space, where we choose to use 180 of the possible words. As discussed previously, signal space saturation, given the choice of a fixed number of labels, is inherently linked to systematicity, such that systematic associations between words and meanings will always produce more saturated signal spaces, since they shrink the space of possible signals (see Figure 5.01).

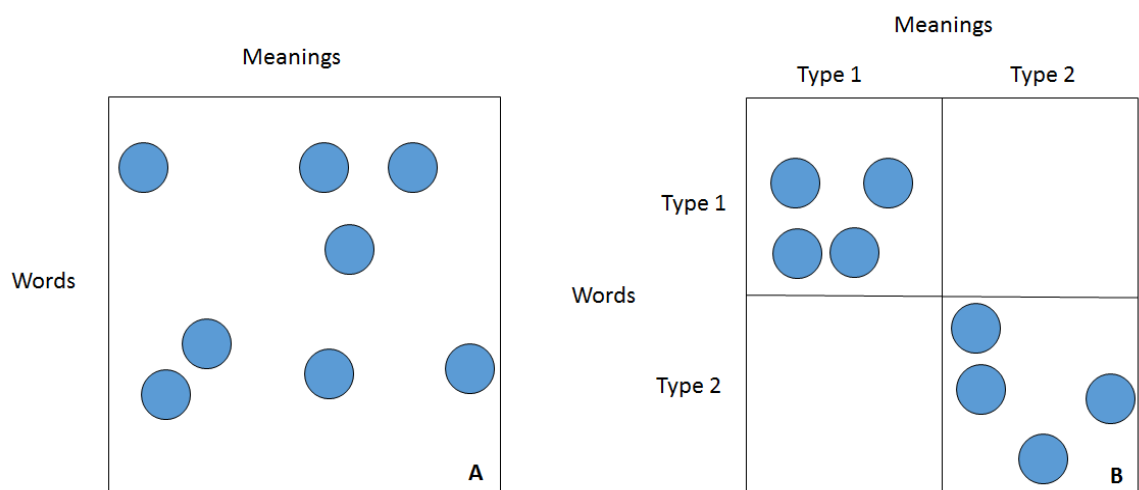


Figure 5.01- A visual representation of the effect of systematicity on signal space saturation. Each blue circle represents a signal-meaning pair and an error term (e.g. arising from

production or perceptual errors) around that word. In figure A, the mapping of words to meanings is unconstrained, and the entire space can be used, whereas under a systematic configuration in figure B, half of the possible signal space becomes unusable and (on average) the error terms around the signal-meaning pairs become closer/eventually overlap.

In Figure 5.01, we can see that the introduction of a systematic mapping between words and meanings necessarily reduces the size of the available signal space, from a large unconstrained area where any type of word can be mapped to any type of meaning, down to two smaller areas. In fact, the visual representation here suggests that this approximately halves the available space, but in our example of cVcV bisyllables, an arbitrary configuration of words-meanings allows for 256 possible labels, whereas a fully systematic one (like that used in Experiment 2 of Chapter 2) gives us 2 much smaller areas each with 16 (2^4) possible labels. Thus, signal space saturation is intimately tied to the number of words required within a signal space, but also to the overall size of that space, with systematicity necessarily infringing on signal space flexibility and resulting in more constrained, highly saturated signal spaces.

5.1.2 Contrastiveness

The concept of contrastiveness is closely and inversely related to signal space saturation. Highly saturated signal spaces will, all other things being equal, result in words that are more similar to one another, and thus, languages that are less

contrastive. However, in Figure 5.01 we can see that within a signal space of a given size, there are multiple ways that words can be assigned that are more or less contrastive. If we compare, for example, type 1 word-meaning mappings in Figure 5.01 to type 2 word-meaning mappings, we can see that the type 1 words are more similar to one another than are the type 2 words, which are more evenly spread out over the possible signal space. Thus, contrastiveness is a measure of how words are chosen **within** a given signal space, whereas signal space saturation is a measure of the total area of a given space that is occupied by word-meaning mappings.

The difference between signal space saturation and contrastiveness is an important one, because without some other process of optimization operating on a signal space, it is possible to have signal spaces with very low levels of saturation that are nevertheless non-contrastive; even arbitrary mappings between words and meanings will sometimes, by chance, be sufficiently similar to one another that they might be confused by virtue of signal similarity alone (e.g. without any constraints we might still arrive by random chance at three words with similar forms and meanings). In fact, a failure to recognize the difference between these two metrics is likely the source of Chapter 2's Experiment 3 failing to find the desired effects: the half-half language used by Monaghan et al. (2011) allowed for a less saturated signal space, and subsequently the labels were chosen in a way that was maximally contrastive. My own labels, on the other hand, were chosen from a less saturated signal space, but chosen randomly, resulting in a larger deviation in the level of

contrastiveness (i.e. on average half-half languages were more contrastive than the fully systematic languages, but few were as contrastive as the languages used in Monaghan et al., 2011).

The prediction arising from my own findings, and those of previous researchers, is that the level of contrastiveness of an artificial lexicon should be proportional to its learnability, especially with regards to the learnability of systematic vs. arbitrary word-meaning mappings.

5.1.3 Previous Findings

The results of the three experiments presented in Chapter 2 of this dissertation closely reflect the findings of previous researchers (i.e. Gasser, 2004; Monaghan et al., 2011) while offering additional insights about the conditions under which systematicity and arbitrariness are favored in artificial language learning contexts. Collectively, the results of those experiments seem to show an overall benefit for learners of systematic languages: the fact that a feature of the word (e.g. their length, in Experiment 2, or their phoneme inventories in Experiments 1 and 3) maps reliably onto a feature of their associated meaning allows participants to learn those configurations of language more readily than participants learning identically sized arbitrary languages. However, this size of the benefit for systematic languages depends on both the type of test (individuation vs. categorization), and the

dimensions along which systematic associations are structured (and thus, the underlying contrastiveness of their labels).

The importance of these two factors is best demonstrated by the difference in learnability between systematic languages in Experiments 1 and 2 presented in chapter 2. For Experiment 2, systematic associations between words and meanings were structured such that monosyllabic words were mapped onto a single type of image (i.e. animals) and trisyllabic words were mapped onto a second type of image (i.e. vehicles); additionally, individual words of each type were chosen in such a way that they were maximally distinctive from one another and thus potentially less confusable. For experiment 1, the word stimuli that I used were much more similar to those used in Monaghan et al. (2011)'s experiments – rather than a category distinction based on number of syllables, the two categories of words differed in their phonology, with the relevant distinction being between sonorant and plosive consonants; additionally, rather than words within those categories being chosen for maximal distinctiveness, they were chosen randomly from that highly constrained signal space. Thus, in terms on contrastiveness, the stimuli used in Experiment 2 were much more contrastive than those used in Experiment 1: between categories the differences between words were larger, and within categories words in experiment 2 were maximally distinct, rather than relatively similar to one another.

The results of these otherwise identical experiments, when compared to each other, are thus illustrative of the degree to which the overall level of contrastiveness influences learnability (see Figure 5.02).

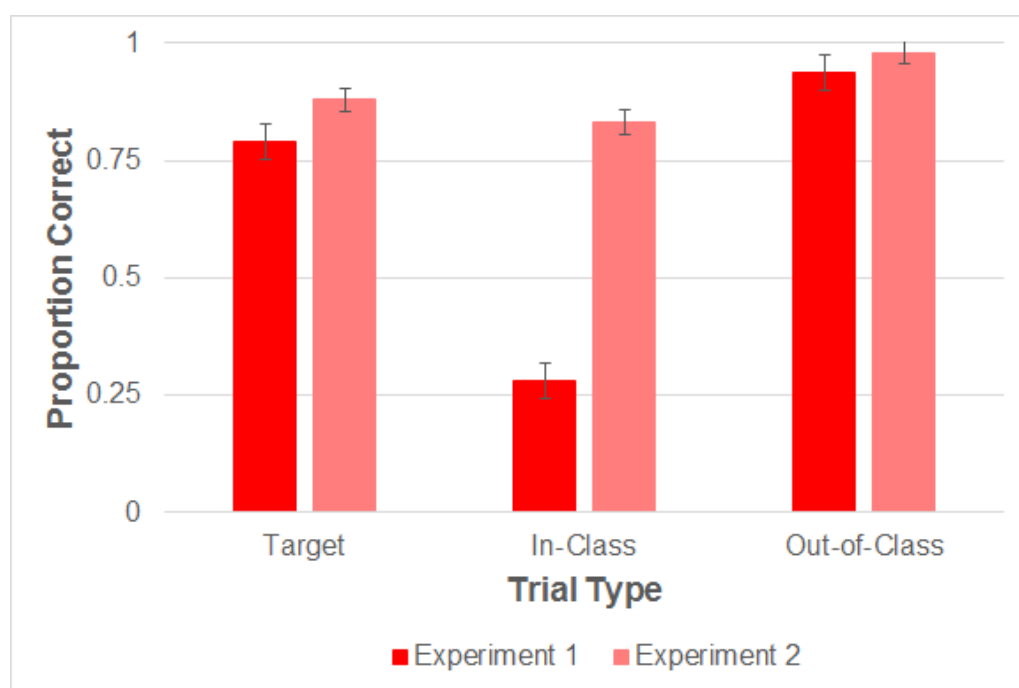


Figure 5.02- Results from Experiments 1 and 2 from Chapter 2 demonstrate that the benefit accrued to systematic language learners is both contingent on the degree of contrastiveness both within and between word types, and also the learnability metric of interest. Error bars show standard error.

The comparison of the results from these two experiments demonstrates, first, that the overall more contrastive languages used in Experiment 2 are substantially easier to learn across the board, regardless of experimental condition. This finding maps nicely onto the general finding of Gasser (2004)’s computational modeling of the

acquisition of artificial lexica: more contrastive languages are, all other things being equal, easier to learn than less contrastive ones.

Additionally, we can see that in general the benefit for systematic language learners accrues primarily in out-of-class distractor trials, and despite the fact that overall performance in the less contrastive experiment 1 is lower, performance on out-of-class trials for systematic language learners is approximately equal and near ceiling: the category distinction is equally apparent for learners of both systematic configurations, and they very rarely make errors on those out-of-class distractor trials, suggesting, at the very least, that they have effectively learned the category structure of their language. Thus, the findings of those two experiments generally support the categorization findings of Monaghan et al.'s experiments, which suggest the same benefit for systematic language learners using an alternative-forced-choice paradigm.

So, the overall comparison of the two experiments suggests a benefit for more contrastive lexica (in alignment with Gasser, 2004), and I replicate previous findings with regards to categorization, but what about individuation? The findings of both Gasser (2004)'s model, and Monaghan et al. (2011)'s models and experiments, suggest that systematic language learners should suffer an individuation penalty, wherein they have difficulty learning to differentiate between individual words within a category. Generally, the findings of both experiments support that

conclusion: performance on in-class distractor trials is worse than performance on target trials for learners of systematic languages (but not for learners of arbitrary languages), suggesting that they are willing to accept in-class distractors at higher-than-expected rates. However the degree to which this is true, and thus, the degree to which systematicity imposes a penalty on individuation varies massively depending on the experiment. In Experiment 2, where individual labels within categories are maximally contrastive, performance on in-class distractor trials is still relatively high: systematic language learners accept in-class distractors at slightly higher rates, but they still seem to be learning to individuate fairly well. In experiment 1 however, this is not the case: performance on in-class distractor trials is well below chance, and is effectively the inverse of performance on target trials, suggesting that participants are unable to differentiate between words within a category and have learned only the structure of the categories that the word-meaning pairs fall into.

5.1.4 Signal space saturation, contrastiveness, and language learning

The results of Experiments 1 and 2 from Chapter 2, combined with the results of Gasser (2004), and Monaghan et al. (2011) suggest that the degree to which benefits and penalties accrue to systematic language learners based on the structure of their signal-meaning mappings is contingent on the degree to which those mappings can remain contrastive in a given signal space. Although the results of Experiment 2 also allowed for some insight into how contrastiveness influences learnability (from a signal space of a given saturation, those configurations of word-meaning pairs that are more contrastive are easier to learn), ultimately the conclusions that could be drawn from the experiments in Chapter 2 were underwhelming. One of the reasons for the results failing to speak to the effect of contrastiveness on learnability has to do with the fact that the manipulation of contrastiveness in those experiments could only be compared between subjects, looking at each artificial lexicon as a whole to determine a contrastiveness value. Additionally, the degree to which learning favored systematic over arbitrary languages (or vice versa) varied as a function of contrastiveness (Experiment 1 vs. Experiment 2), but I could only compare very contrastive lexica to relatively less contrastive ones.

Monaghan et al. (2011)'s suggestion that systematicity and arbitrariness combine in language to give the best of both worlds via a division of labor is an intriguing one, but none of their experimental results or my own speak to the point

at which the benefits of systematicity for learning begin to be outweighed by the costs of confusability due to decreasing contrastiveness. That is, previous results demonstrate that contrastive lexica favor systematicity, and that less contrastive ones favor arbitrariness, but the point at which this switch occurs is unexplored.

Additionally, because previous studies (including my own) involve learning and testing a complete (but small) lexicon, they do not allow for an explanation of how the changing contrastiveness of a single word over time influences its learnability. If, for example, one learns the word ‘*monu*’ for a vehicle, that word is initially the only one of its type, and very contrastive. Subsequently however, one might learn any number of additional similar words: ‘*numo*’, ‘*nonu*’, ‘*muno*’, etc., resulting in a concomitant decrease in the contrastiveness of the previously learned word, which might now be confused with related tokens. To test for this possibility, an experimental protocol that tracks the learnability and contrastiveness of artificial lexica over time is required, as such a protocol will allow not only for an exploration of the effect of overall signal space saturation, but also contrastiveness of individual words at various stages of signal space saturation. Additionally, this type of experimental protocol has the benefit of being more similar to natural language learning, where individual meanings are learned in serial: the fact that research suggests that early acquired portions of the lexicon are more systematic than later acquired portions (Monaghan et al., 2014) makes the exploration of the effect of

changing contrastiveness on learnability even more germane to discussions of language evolution and the structure of the lexicon more broadly.

5.1.5 Sound-Symbolic bootstrapping

One persistent suggestion in the literature is that the learning of sound-symbolic associations between words and meanings are potentially beneficial for language learning. To wit, Imai & Kita (2014) suggest that “...recent findings from cognitive psychology, cognitive neuroscience, and developmental psychology, cognitive and anthropological linguistics converge on the view that iconicity plays a core role for philogenesis and ontogenesis of language...”.

The proposal that motivated associations between words and meanings are learned more readily than arbitrary ones is, as previously discussed, fairly well established in both the experimental literature (e.g. Nielsen & Rendall, 2012; Nygaard, Cook, & Namy, 2009; Perniss & Vigliocco, 2014) as well as corpus analyses demonstrating that sound-symbolic associations are manifest both in language broadly (Blasi et al., 2015) and specifically in early acquired parts of the lexicon (Monaghan et al., 2014) and child-directed speech (Akita, 2011; Ogura, 2006). However, the idea that sound-symbolism, and motivated associations between words and meanings more generally, bootstraps the acquisition of language more broadly relies on the suggestion that learning motivated word-meaning associations

influences learning over-and-above the enhancement provided to the learnability of individual tokens.

Unfortunately, until recently suggestions that motivated associations between words and meanings bootstrap the acquisition of other parts of the lexicon haven't been made entirely clear (e.g. Nielsen, 2011; Cuskley, 2013; Ramachandran & Hubbard, 2001). Recently, other researchers have begun to center in on more fully explicated bootstrapping hypotheses: for example, Imai & Kita (2014) have suggested that motivated associations help establish reference and lexical representation, both because the motivatedness of tokens makes some word-meaning pairs more salient, and also because this salience allows infants to extract relevant features from complex visual scenes (see also Perniss & Vigliocco, 2014). Here, I will refer to these types of bootstrapping arguments as *referential bootstrapping*, which I discussed in the introduction: motivated connections between words and meanings function to allow infants to establish reference and learn to attach linguistic sounds to meanings in the environment.

In addition to referential bootstrapping, in the introduction I also introduced the idea of *conceptual bootstrapping*, which suggests that the combination of systematicity and motivatedness might allow naïve learners to more easily pick up on category distinctions that are relevant to their specific language and establish concepts that can be relevant for the learning of subsequent arbitrary tokens. This

type of bootstrapping argument suggests not only that motivatedness and systematicity might highlight salient dimensions along which categories are structured, but also that this ability might be leveraged to establish the existence of concepts and categories generally in the mind of the naïve learner.

I will call a final type of bootstrapping that might be of interest *simple bootstrapping*. This version of bootstrapping merely suggests that the learning of motivated word-meaning mappings increases the subsequent learnability of arbitrary tokens by some unspecified mechanism: this is the unspecified version of bootstrapping that we might ascribe to previous authors (myself included) who were unclear about what bootstrapping explanation they favored, and instead invoked the concept of bootstrapping non-specifically. Although this unspecified bootstrapping account is untenable, because it lacks a mechanism, in this chapter I explore it as a possibility because it is experimentally approachable.

Unfortunately, experiments using adult participants who presumably have already learned to establish lexical reference cannot test the tenability of the referential bootstrapping hypothesis. Similarly, conceptual bootstrapping might be difficult to test with adult participants: in the first case, it is obvious that they have already established concepts and categories generally, and in the second case, the kind of stimuli that are typically used in these experiments belong to easily recognizable categories. Testing the simple bootstrapping hypothesis however is

much simpler, requiring only a temporal dimension to learning where participants are taught an initially motivated language and then tested for their ability to learn subsequent arbitrary word-meaning mappings. For this reason, I test the simple bootstrapping hypothesis here, not because it is an explanation that merits serious consideration, but because exploring it might tell us about the feasibility of more well-stated and plausible bootstrapping hypotheses.

5.2 Investigating signal space saturation and sound-symbolic bootstrapping

It is possible to test the influence that changing contrastiveness has on learnability of systematic languages, and the simple bootstrapping hypothesis for the benefit of motivated languages simultaneously. In Chapter 4 I demonstrated that, other than an early benefit for motivated languages, motivated systematic and non-motivated systematic lexica are approximately equally easy to learn; thus, here I can simultaneously use both while maintaining an ability to compare the results broadly to those in Chapters 2 and 3 of this dissertation.

5.3 Experiment 7

To explore both the effect of changing contrastiveness on language learning and the simple bootstrapping hypothesis, an experimental protocol was required that met a number of criteria. To test the simple bootstrapping hypothesis, an experimental design was required that separated learning and testing such that the motivatedness

of learned word-meaning mappings could change over time. Simultaneously, exploring contrastiveness more fully required measurement of learnability over multiple rounds of training and testing.

Specifically, a test of the bootstrapping hypothesis required that participants could be taught an initial motivated language, then transition at some point to learning arbitrary word-meaning mappings. Crucially however, the learnability of later-acquired arbitrary labels needed to be compared to some sensible baseline, such that any difference in learnability of early acquired vs. late acquired words could be traced to bootstrapping, rather than some other factor (primacy and recency effects, reduced performance due to increasing cognitive demand, etc.).

In order to test the effect of contrastiveness on learnability, we need a way to quantify how the introduction of additional words impacts on the signal space saturation and the contrastiveness of existing labels. In previous experiments (i.e. Experiment 3 from Chapter 2), the contrastiveness metric used failed to capture the influence that the presence of additional similar words had on learnability, and as such a new contrastiveness metric was required that could be calculated for words in this new experimental protocol. The new contrastiveness metric used for this experiment, in addition to changing over the process of learning, should also straightforwardly predict the learnability of word-meaning pairs over time, rather

than any deficit in learnability simply being traceable to participants being required to learn additional word-meaning pairs over time.

Given these requirements, the simple bootstrapping hypothesis suggests that participants who learn motivated word-meaning mappings early before switching to learn arbitrary word-meaning mappings later should perform significantly better than participants who do not have the benefit of this scaffolding. In addition to this prediction, we should find that lexica that are more contrastive will be easier to learn, and that for individual words, performance will fall off over time as a function of the decreasing contrastiveness of words to one another as additional word-meaning mappings are learned, and that this penalty should be especially prevalent for learners of systematic lexica due to their lower levels of contrastiveness.

5.3.1 Methods

Participants

Participants were 49 (32 female) students and members of the general population recruited from the University of Edinburgh ($n = 31$) and the University of Lethbridge ($n = 18$) subject pools, and were compensated 3.00 GBP and \$5.00 CAD respectively for their participation in the experiment, which took approximately 25 minutes. All participants were fluent English speakers between 17 and 35 years of age (Mean = 21.92, StDev = 3.81) with normal hearing and normal or corrected-to-

normal vision. Each of these experimental participants was assigned randomly to one of the four experimental conditions. Ethical approval was obtained locally at both the University of Edinburgh and the University of Lethbridge, adhering to both British Psychological Association and American Psychological Association guidelines, and informed consent was obtained from all participants.

Experimental Design

The experimental protocol created for this experiment differed from those of previous experiments in that participants were exposed to multiple rounds of training and testing, learning a complete artificial lexicon over the course of these multiple bouts rather than all at once. Specifically, participants were trained with an initial language, and then exposed to alternating rounds of testing and training where they learned new words in each training round, and then were tested on those new words in addition to all previously learned words in each testing round.

This manipulation allowed me to simultaneously explore the simple bootstrapping hypothesis and the effect of changing contrastiveness on learnability. Specifically, participants in the experiment were split into 4 conditions in a 2 (Early language: Motivated or Arbitrary) \times 2 (Late language: Motivated or Arbitrary) design, as shown below in Table 5.01.

<i>Condition</i>	<i>Early Language (First 8 Labels)</i>	<i>Late Language (last 8 labels)</i>
<i>1</i>	<i>Motivated Systematic</i>	<i>Motivated Systematic</i>
<i>2</i>	<i>Motivated Systematic</i>	<i>Arbitrary</i>
<i>3</i>	<i>Arbitrary</i>	<i>Arbitrary</i>
<i>4</i>	<i>Arbitrary</i>	<i>Motivated Systematic</i>

Table 5.01- *The four conditions for used in the growing lexicon experiment. Participants learned an initial language that was either arbitrary or motivated and systematic, then later learned additional tokens that were either arbitrary or motivated and systematic, in a 2x2 factorial design.*

Label Stimuli

All words created for this experiment were bisyllables in cVcV order.

To create labels that were either motivated and systematic or arbitrary I used a total of 8 possible consonants. Following the stimuli used in Experiment 1 of this dissertation I used the plosive consonants /t/ and /p/ in contrast with the consonants /m/ and /n/ for sound symbolic labels. For the creation of arbitrary labels I needed two pairs of consonants that I was reasonably certain either a) wouldn't have any associated sound symbolic bias and/or b) would have equal small biases. Thus, I selected four unvoiced fricative consonants, contrasting /s/ and /f/ with /θ/ and /ʃ/ (see Table 5.02).

<i>Sound Symbolic</i>		<i>Arbitrary</i>	
<i>Plosive</i>	<i>p, t</i>	<i>Type 1</i>	<i>f, s</i>
<i>Sonorant</i>	<i>m, n</i>	<i>Type 2</i>	<i>ʃ, θ</i>

Table 5.02- Consonants used for the construction of words in Experiment 7.

Based on these pairs of consonants, 8 possible words were created for each of the word types (as that is the maximum number that could be used for any one participant) for each participant.

To create the 8 possible words of each type, 8 consonant skeletons were selected from 16 possible skeletons based on the available consonants for that type. For example, given the consonants /t/ and /p/, there are four possible consonant configurations (t_t_, t_p_, p_t_, and p_p_). I thus created four copies of each of the possible configurations, for 16 total possibilities, then selected 8 consonant skeletons randomly from this total of 16 (thus, each participant would on average have two labels with each consonant configuration, but could have anywhere from 0 to 4 of each type).

Thus, we might arrive at these possible consonant configurations for a participant. I created complete words from these consonant configurations by assigning the four possible vowels semi-randomly with the constraint that no duplicate words were created and that each of the four possible vowels occurred an equal number of times in both the first and second syllable (Table 5.03).

Plosive Labels	Arbitrary Labels Type 1
tato	ʃʌθʌ
tepe	θaθo
pota	ʃeθe
tete	θeθe
tʌta	ʃoʃʌ
popʌ	θʌʃa
tʌpo	θoθa
papʌ	θaʃo
Sonorant Labels	Arbitrary Labels Type 2
nʌna	fosa
namo	sese
momo	fʌsʌ
nʌnʌ	safo
mone	sasa
memʌ	sesʌ
mema	fofe
name	sʌso

Table 5.03- *An example of the possible set of labels for an experimental participant from which lexica were chosen.*

Once these possible labels were created for each participant, the appropriate numbers of labels were randomly selected based on experimental condition, as below:

Motivated Systematic early- Motivated Systematic late- 8 Plosive Labels, 8

Sonorant Labels

Motivated Systematic early- Arbitrary late- 4 Plosive Labels, 4 Sonorant Labels, 8

chosen randomly from the arbitrary labels

Arbitrary early- Arbitrary late- 16 arbitrary labels

Arbitrary early- Motivated Systematic late 8 chosen randomly from the arbitrary labels, 4 plosive labels, 4 sonorant labels

This word generation procedure allows for a subset of 16 words for each participant to be selected from 256 possible words ($4 \times (2 \times 4 \times 2 \times 4)$) in such a way that I ensure that the overall use of vowels is unbiased while allowing for a range of possible consonant configurations that ensures that individual participants will have languages that are more or less contrastive within word types. Words were presented to participants as both auditory stimuli (see below) and also on screen in the orthographic form with the following substitutions (Table 5.04):

<i>Consonants</i>		<i>Vowels</i>	
<i>IPA</i>	<i>Orthographic</i>	<i>IPA</i>	<i>Orthographic</i>
<i>p</i>	<i>p</i>	<i>a</i>	<i>ah</i>
<i>t</i>	<i>t</i>	<i>o</i>	<i>oh</i>
<i>m</i>	<i>m</i>	<i>e</i>	<i>ay</i>
<i>n</i>	<i>n</i>	<i>ʌ</i>	<i>uh</i>
<i>s</i>	<i>s</i>		
<i>f</i>	<i>f</i>		
<i>ʃ</i>	<i>sh</i>		
<i>θ</i>	<i>th</i>		

Table 5.04- *A representation of the IPA symbol and associated orthographic form presented to participants in Experiment 7. Auditory stimuli were generated as closely as possible to their IPA notation, but orthographically labels were presented in such a way that the orthography would be more accessible to IPA-naïve experimental participants.*

Each of the 256 possible total words was created as an audio file using Apple Talk with the Victoria voice (As in Experiment 1 of Chapter 2). Because apple talk does not use phonetic symbols, the phonetic representations given for each of the phonemes (especially the vowels) is inexact, although I ensured that the auditory representation was as close to the intended sequence as possible.

Image Stimuli

The image stimuli used in this experiment were created using a radially constrained mathematical formula which created pairs of curved and jagged image forms from the same set of randomly generated calculus points. The resulting image pairs were identical to one another except in the curvature of their lines. Full details of the image generation technique are provided in Nielsen and Rendall (2011). Using this methodology I created a large set of 192 total images (96 total pairs) from which I selected 12 rounded and 12 jagged images for their distinctiveness. No matching images from any seed were chosen.

For each participant, 8 of each of these 12 possible image types was used, for a total of 16 images for each experimental participant (8 curved, 8 jagged).

Procedure

In this experiment participants were taught associations between pseudowords and meanings (images of either jagged or curved shapes) over the course of alternating rounds of training and testing. In each round of training, they were exposed to a set of new word-meaning pairs a total of 6 times each in randomized order. After training, they were tested on their ability to remember the correct image for each of the pseudowords that they have learned, both in the immediately preceding training

block and in all previous rounds, by clicking on the correct image from a field of possible images.

Training

In each block of training, participants were exposed to a new set of paired pseudowords and images a total of 6 times each in randomized order. In the first training block, participants learn 4 pseudowords, whereas in the subsequent 6 blocks they learn only 2 new pseudowords per block.

On each training exposure, participants were shown a fixation cross, followed by the appropriate image appearing on screen. After a delay of 500 ms, the orthographic representation of the label was displayed to the participant below the image. One second later, the label was played to them auditorily via headphones. After another 2 seconds the label was played for them a second time. Finally, after a final 2 second delay, the fixation cross came back up to signal the start of a new training exposure.

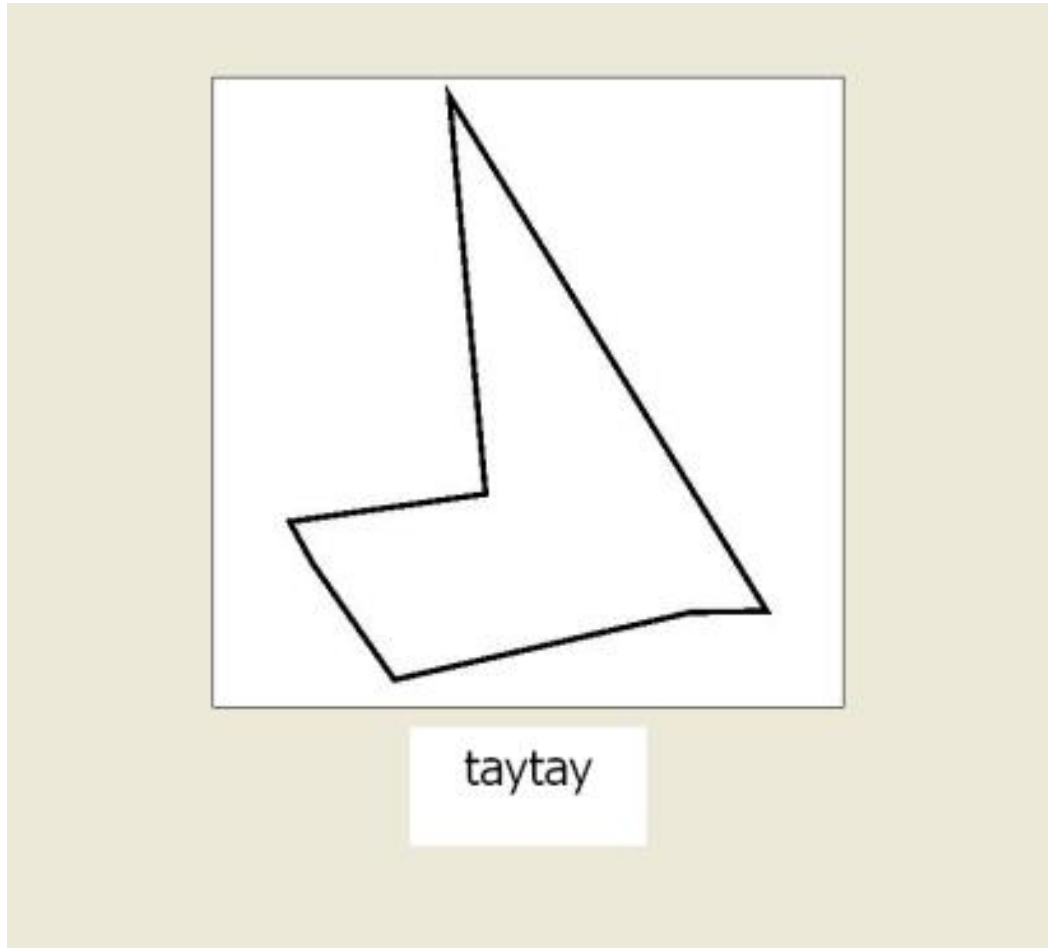


Figure 5.03- Example training trial from Experiment 7 showing a participant being exposed to a jagged image with a plosive label. In addition to the label being presented to the participant in orthographic form, it was also presented to them via headphones.

Testing

In each block of testing, participants were tested for their ability to correctly pair pseudowords with their appropriate image. On each trial participants were shown a field of between 4 and 16 possible images (all of the images that they had seen up to

that point in the experiment) along with a single pseudoword, that was presented to them both orthographically in the center of the screen and auditorily via headphones (Figure 5.04).

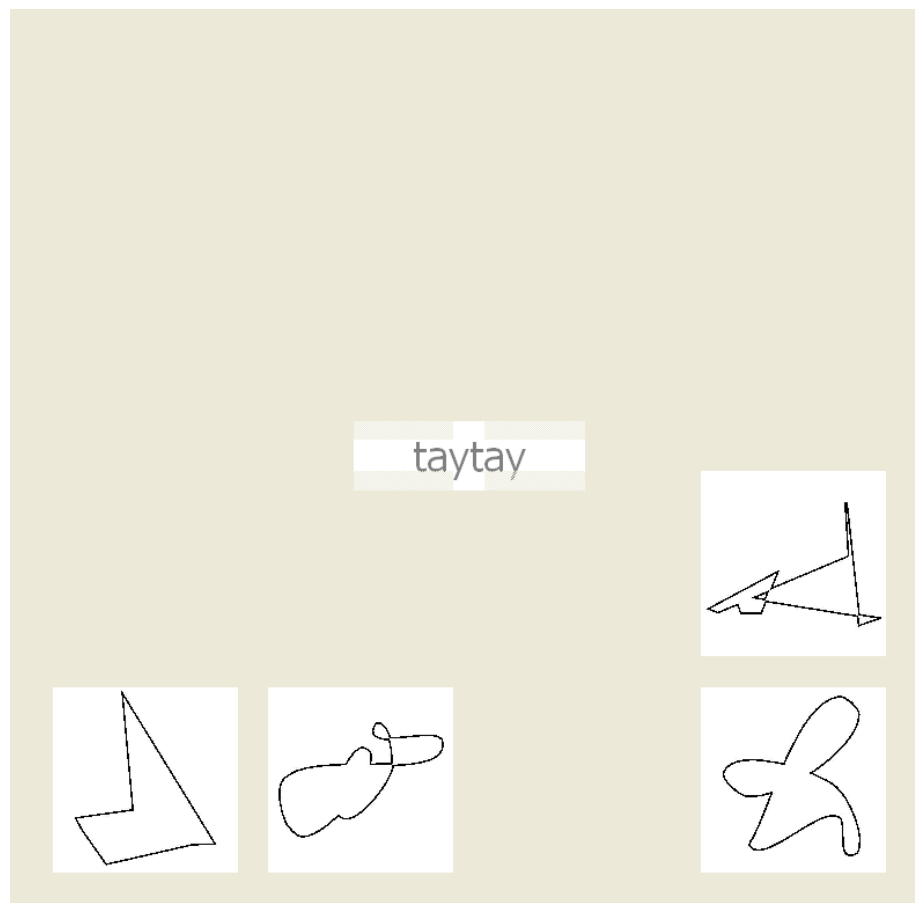


Figure 5.04- Example of a testing trial for a participant in Experiment 1. The participant is presented with the label 'taytay' both in the orthographic form shown on screen and via headphones, then tasked with choosing the correct meaning for that word from the available options.

Participants made their selection on each trial by clicking on their choice of appropriate image for a given label, which recorded their response and progressed them to the next trial.

In each testing block, the location of the possible images to be selected was randomized to ensure that participants were learning associations between words and shapes, rather than words and response locations.

As outlined above, Experiment 7 used a 2x2 factorial design, with participants learning either a motivated systematic or arbitrary initial language, then later learning either a motivated systematic or arbitrary late language. The early language consisted of the first 8 pairs of words and meanings learned over the first 3 rounds of training and testing, while the late-acquired language consisted of the remaining 8 pairs of words and meanings learned over the final 4 rounds of training and testing (see Table 5.05).

<i>Round</i>	<i># Trained</i>	<i># Tested</i>	<i>MS Early MS Late</i>	<i>MS Early Arb Late</i>	<i>Arb Early Arb Late</i>	<i>Arb Early MS Late</i>
<i>1</i>	<i>4</i>	<i>4</i>	<i>4 MS</i>	<i>4 MS</i>	<i>4 Arb</i>	<i>4 Arb</i>
<i>2</i>	<i>2</i>	<i>6</i>	<i>2 MS</i>	<i>2 MS</i>	<i>2 Arb</i>	<i>2 Arb</i>
<i>3</i>	<i>2</i>	<i>8</i>	<i>2MS</i>	<i>2MS</i>	<i>2 Arb</i>	<i>2 Arb</i>
<i>4</i>	<i>2</i>	<i>10</i>	<i>2MS</i>	<i>2 Arb</i>	<i>2 Arb</i>	<i>2MS</i>
<i>5</i>	<i>2</i>	<i>12</i>	<i>2MS</i>	<i>2 Arb</i>	<i>2 Arb</i>	<i>2MS</i>
<i>6</i>	<i>2</i>	<i>14</i>	<i>2MS</i>	<i>2 Arb</i>	<i>2 Arb</i>	<i>2MS</i>
<i>7</i>	<i>2</i>	<i>16</i>	<i>2MS</i>	<i>2 Arb</i>	<i>2 Arb</i>	<i>2MS</i>
<i>Total</i>	<i>16</i>	<i>70</i>	<i>16 MS</i>	<i>8 MS, 8 Arb</i>	<i>16 Arb</i>	<i>8 Arb, 8 MS</i>

Table 5.05- *The number of new words trained and number of words tested at each experimental block in Experiment 7. Additionally, a description of the structure of experimental blocks for participants in each of the four conditions of the 2x2 factorial design used in Experiment (MS= Motivated Systematic, Arb=Arbitrary).*

Contrastiveness and Confusability

The contrastiveness metric used in Experiment 3 of Chapter 1 of this dissertation, and also by Monaghan et al. (2011) is ultimately not one that is appropriate for exploring the types of questions that Experiment 7 here seeks to answer.

Specifically, the average edit distance of a single word to all other words of its type fails to adequately capture the degree to which that label is contrastive from other labels of its type. As a simple demonstration, consider the example presented below in Table 5.06.

<i>Label of Interest</i>	<i>In-Class Label</i>	<i>Edit Distance</i>	<i>Exposure Round</i>	<i>Average Edit Distance</i>
<i>naymoh</i>	<i>mohmay</i>	<i>3</i>	<i>1</i>	<i>3</i>
	<i>nuhmoh</i>	<i>1</i>	<i>2</i>	<i>2</i>
	<i>maymay</i>	<i>2</i>	<i>3</i>	<i>2</i>
	<i>muhnah</i>	<i>4</i>	<i>4</i>	<i>2.5</i>

Table 5.06- *An example of the comparison of a given label ('ne mo') to related labels, and the effect that the introduction of those new labels has on the average edit distance- my previously used contrastiveness metric.*

The example language shown in table 5.06 demonstrates quite clearly why average edit distance fails as a metric of contrastiveness when looking at individual labels: At first, the introduction of more labels decreases the average edit distance, which seems to capture contrastiveness, but in exposure round 4 a maximally different label is introduced and the average edit distance actually becomes higher. In one sense, this seems reasonable: 'muhnah' is maximally different from 'naymoh', and thus the average distance between the labels increases, but the introduction of 'muhnah', which is maximally different, should not result in the suggestion that 'nuhmoh' would suddenly be less easily confused with 'muhnah'. Comparing the contrastiveness of single words to each other using their hamming distance is appropriate, as we saw in Chapter 3, as is averaging the contrastiveness of every word in a language to each other ('muhnah' is not similar to our target word, but is quite similar to other words in the language), as we saw in Experiment 1 of Chapter

2. However, when looking at a single target word a contrastiveness value based on the similarity to all other words (but not those words to each other) fails to capture similarity adequately- the introduction of a new dissimilar label word should not make an existing similar label easier to learn. No matter what new label is introduced, and no matter the similarity to the label of interest, the requirement to learn an additional new token should never result in the prediction that performance will actually improve. Thus, a new metric was required that would capture not only the fact that the introduction of additional labels should always reduce contrastiveness, but that similar labels should reduce contrastiveness more.

To capture these effects, I created a new metric for contrastiveness designed to predict the possibility that participants would confuse the word of interest with any other previously learned labels. This metric is thus a measure of confusability, with low values suggesting a lower probability that participants will confuse labels in their language (thus, a low value = a more contrastive language). Confusability is the inverse of the edit distance between two tokens (so, the labels ‘naymoh’ and ‘mohmay’, which have an edit distance of 3, have a confusability value of 1/3). For a given word, the overall confusability value is calculated by summing the confusability values from the comparisons of that word with all other words of the same type, as seen in table 5.07.

<i>Target Label</i>	<i>Comparison Label</i>	<i>Edit Distance</i>	<i>Exposure Round</i>	<i>Avg. Edit Distance</i>	<i>Confusability</i>	<i>Summed Confusability</i>
<i>naymoh</i>	<i>mohmay</i>	<i>3</i>	<i>1</i>	<i>3</i>	<i>1/3</i>	<i>.33</i>
	<i>nuhmoh</i>	<i>1</i>	<i>2</i>	<i>2</i>	<i>1</i>	<i>1.33</i>
	<i>maymay</i>	<i>2</i>	<i>3</i>	<i>2</i>	<i>½</i>	<i>1.83</i>
	<i>muhnah</i>	<i>4</i>	<i>4</i>	<i>2.5</i>	<i>¼</i>	<i>2.08</i>

Table 5.07- An example of the calculation of the new confusability metric used in Experiment 7.

Although I use this specific confusability metric (inverse edit distance) in the analyses presented in this chapter, the general findings hold under a number of models of confusability, shown in Table 5.08.

<i>Edit Distance</i>	<i>Contrastiveness Metric</i>			
	<i>Contrast A</i>	<i>Contrast B</i>	<i>Contrast C</i>	<i>Contrast D</i>
<i>1</i>	<i>1</i>	<i>1</i>	<i>1</i>	<i>1</i>
<i>2</i>	<i>1</i>	<i>0</i>	<i>1/2</i>	<i>1/4</i>
<i>3</i>	<i>1</i>	<i>0</i>	<i>1/3</i>	<i>1/9</i>
<i>4</i>	<i>1</i>	<i>0</i>	<i>1/4</i>	<i>1/16</i>

Table 5.08- Calculation rules for 4 possible contrastiveness metrics tested for Experiment 7. In Metric A, all labels of the same type are weighed evenly, and assumed to be equally confusing. In Metric B, only neighbours with an edit distance of 1 influence contrastiveness/confusability. Metric C reflects the metric used in this chapter, where additional labels add the inverse of their edit distance to the summed contrastiveness metric. Metric D works similarly, but using an inverse square law.

This new confusability metric has a number of benefits. First, early learned words will have relatively low values for their summed confusability, but their summed

confusability will be higher every time they are subsequently tested as a function of the number of additional words learned and their similarity to those words- thus, performance of early acquired words can be compared across subsequent testing blocks, allowing me to test whether the introduction of additional confusable labels can impair performance on previously learned labels. In addition to an exploration of the effect of increasing confusability / decreasing contrastiveness, this metric assures that later learned words will have higher confusability.

It bears noting, in addition to the strengths of the metric, that it is not truly a **measure** of confusability: I measure how often language learners confused words based on performance in the experiment. The confusability metric I use is, rather than being a measure, a predictor based on how I suspect that the presence of additional similar words might affect learnability.

Other metrics of contrastiveness and confusability

It is worth mentioning that there are multiple literatures outside of research about motivatedness and systematicity that have their own interpretations of contrastiveness and confusability and have developed metrics to quantify those features. Although I do not make use of those metrics here, it is important to acknowledge them and link them to my own notions about contrastiveness and confusability. Ultimately, the deployability of some of these earlier-described

metrics for confusability are inadequate for my present purposes for a number of reasons, which I will outline in discussing them below.

Historically, Miller & Nicely (1955) were amongst the first to explore the perceptual confusability of (English) consonants systematically in a way that allowed them to make general statements about confusability, rather than simply cataloguing perceptual errors. Miller & Nicely created confusion matrices for each of the 16 most common consonants in English, such that the proportion of the time that a given consonant was either recognized, or mistaken for another of the consonants of interest, was recorded. These confusion matrices were calculated both with neutral vocal stimuli and under a range of noise conditions. The overall confusion matrices under various levels of noise calculated by Miller & Nicely can be compressed such that they are slightly less burdensome (one can, for example, focus on how much confusion there is between relatively more or less similar phonemes), or considered in terms of some of the phonological features of the studied phonemes.

Aside from its mathematical complexity, the applicability of Miller & Nicely (1955)'s classic confusion data to the work presented here is limited by a number of factors. First, Miller & Nicely consider only consonant phonemes, which means that for the present study we would have to ignore confusability caused by vowel similarity. Second, the majority of the data from Miller & Nicely deals with the confusability of phonemes under various types of noise, which is not ideal for the present study for at

least two reasons: a) that stimuli are presented in such a way that noise is limited, and b) that it's difficult to determine which confusability matrix is the most appropriate for use under these conditions. Third, Miller & Nicely's data is framed entirely in terms of perceptual confusability (i.e. what a participant heard), whereas we are interested in confusability more generally (which might include perceptual confusability, but also memorability, and potentially even productive confusability). Finally, Miller & Nicely's notion of confusability, in focusing on individual phonemes, has limited applicability to the confusability of entire word forms.

Following Miller & Nicely (1955), a number of other researchers tackled some of those shortcomings and extended the notion of confusability based on similar measurements. Wicklegren (1965, 66) found for example that phoneme similarity also influenced confusability in short term memory, while others (e.g. Bailey & Hahn, 2001) demonstrated that phoneme similarity played a role in determining the confusability of whole words. In 2006, Bailey & Hahn returned to the issue of similarity in an attempt to answer two questions that are also directly applicable to the work that I present here: i) Is there a single notion of "phoneme similarity" that underlies perceptual, memory, and other observed differences, and ii) What is the best measure of phoneme similarity?

To tackle these questions, Bailey & Hahn (2006) compared confusability metrics based on phonological features (like manner + place of articulation) to those based on confusability of perception (e.g. Miller & Nicely, 1955), production (e.g. Dell & Reich, 1981), and short-term memory retrieval (e.g. Wicklegren, 1965). For phonological feature metrics, Bailey & Hahn explored both S_{PMV} , which is a similarity metric based on place and manner of articulation + voicing, and the natural class metric from Frisch (1996). For confusability metrics, Bailey & Hahn used the phoneme confusability in short term memory from Wicklegren (1966), the speech production phoneme confusability from Shattuck-Hufnagel & Klatt (1979), and a measure of perceptual confusability from Luce (1986) under six signal-to-noise ratios.

Bailey & Hahn (2006) found that S_{PMV} , the simplest featural metric was the best predictor of confusability. Further, they found through a comparison of the various metrics to one another that psychological estimates of the differences between phonemes were better predictors than estimates of their commonality. Of note, they also find that the confusability relationship is non-linear – that is, more dissimilar phonemes are confused more easily, but not to the degree one would expect based on the confusability where only a single feature differs.

Bailey & Hahn's findings offer a number of potential insights into the confusability metric that we use here, although the benefit of using S_{PMV} or some other,

potentially more robust, featural metric are not entirely clear. Bailey & Hahn's findings improve over those of Miller & Nicely (1955) for our purposes in that they take more than perceptual confusability into account, but they are still formulated such that they refer to differences between phonemes, and not an aggregate predictor for the similarity or dissimilarity of entire words. Computing such a value from Bailey & Hahn's data may be possible, but would require additional experimental motivation in determining how to weigh the influence of various phonemes relative to one another.

Other authors, like Nowak & Krakauer (1999) have considered perceptual confusability more abstractly and even more mathematically than Miller & Nicely (1955). Using computational models, they demonstrate that when there is a fitness payoff, languages can evolve such that the sounds are selected to minimize their similarity. These model languages can add new sounds to increase the number of describable objects, but this only increases confusability. According to Nowak & Krakauer then, the process of combining discrete sounds into words is a direct response to the pressure for expressivity (which they call unlimited semantic representation). Under this system, the authors suggest that word recognition is based on identification of each individual phoneme in the word. Although they do not discuss the underlying representation of these phonemes, this suggests that phoneme similarity, whatever its underlying metric, should have important implications for the confusability of words to one another, although, again, a fitness

pressure operating on the lexicon should select for configurations of words that are maximally distinctive.

Collectively, the approaches outlined above suggest that predicting confusability is both mathematically and practically quite complex. Featural similarity models are theoretically motivated, and do a fairly good job of predicting actual confusability in a wide range of experimental manipulations, which suggests that those features bear some psychological resemblance to the perceptual and cognitive features relevant to human language learners. Despite this fact, our own research suggests that phonological feature encodings do not actually predict the ability of human participants to learn artificial languages (Chapter 3), which casts some doubt on the applicability of the metrics described above for the research presented here.

Certainly, taking advantage of more robust and grounded (both ecologically and theoretically) metrics like those developed by Bailey & Hahn (2006) has its benefits, but those benefits must be weighed against practical issues as well. First, the metrics outlined in previous research are often quite opaque, especially to those who are not proficient with some fairly complicated matrix algebra (myself included). Relatedly, the metrics are not parameterized in an accessible way that would make them useful for applying to new experimental manipulations. Third, the metrics that I have discussed are generally based on differences between phonemes in equivalent locations in otherwise identical words, rather than a comparison of the confusability of entire words to one another. Finally, and relatedly, extending these methods

directly (even if it were easy to do so) to comparisons of whole words might be difficult, as weighing the relative importance of similarity at different loci in a word could be problematic without further empirical support.

This is, of course, not to be a naysayer entirely about previously established methods for calculating or predicting confusability. Even if we were to find that the metric used here was superior to previously described ones, the insights from exploring those methods can provide insight for how my confusability metric fits into the broader psycholinguistic literature. In future, extensions of the work presented in this chapter should consider incorporating these considerations, if not the actual confusability metrics developed by others, more completely. For the present study however, the relative simplicity of the confusability predictor that I have created is a strength in that it is easily approachable and makes simple predictions in line with the findings of experiments presented earlier in this dissertation. Ultimately, I hope that the continuation of work like this by myself and others can at the very least compare this type of confusability predictor to the performance of other predictors, but as an experiment that is the first of its kind I have here favored simplicity and ease-of-interpretation over other factors.

Data Analysis

Responses for each trial of the experiment were coded for correctness and then analysed using a logistic mixed effects analysis of the relationship between correctness and experimental condition. The analysis was conducted using R (R Core Team, 2012) and lme4 (Bates et al. 2015). As a test of my confusability metric, differences between the four experimental conditions on early learned words in the early trials and later performance on those early learned words after more words were introduced and confusability increased was the most relevant comparison. As a test of the bootstrapping hypothesis, the differences in performance between conditions comparing early-learned words tested late and later-learned words tested late was the most relevant comparison; if the bootstrapping hypothesis is correct, then for MS Early -> Arb Late languages performance on late acquired words (i.e. arbitrary word-meaning mappings) should be better than either performance on early acquired (i.e. Motivated Systematic) words tested later, late acquired arbitrary words acquired in other conditions (Arb Early -> Arb Late), or both. To best explore these possibilities, I combined Exposure round (early vs. late) and testing block (early vs. late) into a single factor: Trial Type, which is shown Table 5.09.

	<i>Early Acquired</i>	<i>Early Acquired</i>	<i>Late Acquired</i>
	<i>Early Tested</i>	<i>Late Tested</i>	<i>Late Tested</i>
<i>Rounds Trained</i>	<i>1-3</i>	<i>1-3</i>	<i>4-7</i>
<i>Rounds Tested</i>	<i>1-3</i>	<i>4-7</i>	<i>4-7</i>

Table 5.09- *Trial Type (Exposure Block x Testing Block) factor used for data analysis in Experiment 7.*

For the logistic mixed effects analysis of the relationship between correctness and experimental condition I used Trial Type and Condition (and their interactions) as fixed effects, with subject as a random effect. P-values for fixed effects and their interaction were obtained using likelihood ratio tests of the full model compared against the model without the effect; other p-values reported below were obtained via the normal approximation.

A second analysis, identical to the first but including my new contrastiveness metric (and the associated interactions) as a fixed effect was also conducted to explore the degree to which contrastiveness influenced the relative learnability of systematic vs. arbitrary word-meaning associations.

The simple bootstrapping hypothesis suggests that late-learned arbitrary word-meaning mappings should be easier for participants who have previously learned a systematic language than for participants who have learned an arbitrary early language. To test this possibility, I performed a two-sample t-test on

comparing performance on these late trials between these two groups. Additionally, I compared performance on late-learned motivated systematic associations between participants who had learned either motivated or arbitrary associations in the early acquired portion of their lexicon.

Finally, I conducted a planned analysis comparing performance on block 1 between motivated systematic language learners and arbitrary language learners as an additional test of the early effect of motivatedness found in Chapter 4.

5.3.2 Results

Logistic Mixed Effects Regression I

Performance across the three levels of Trial Type (Early Acquired- Early Tested; Early Acquired- Late Tested; and Late Acquired- Late Tested) for each of the four experimental conditions is shown in Figure 5.05.

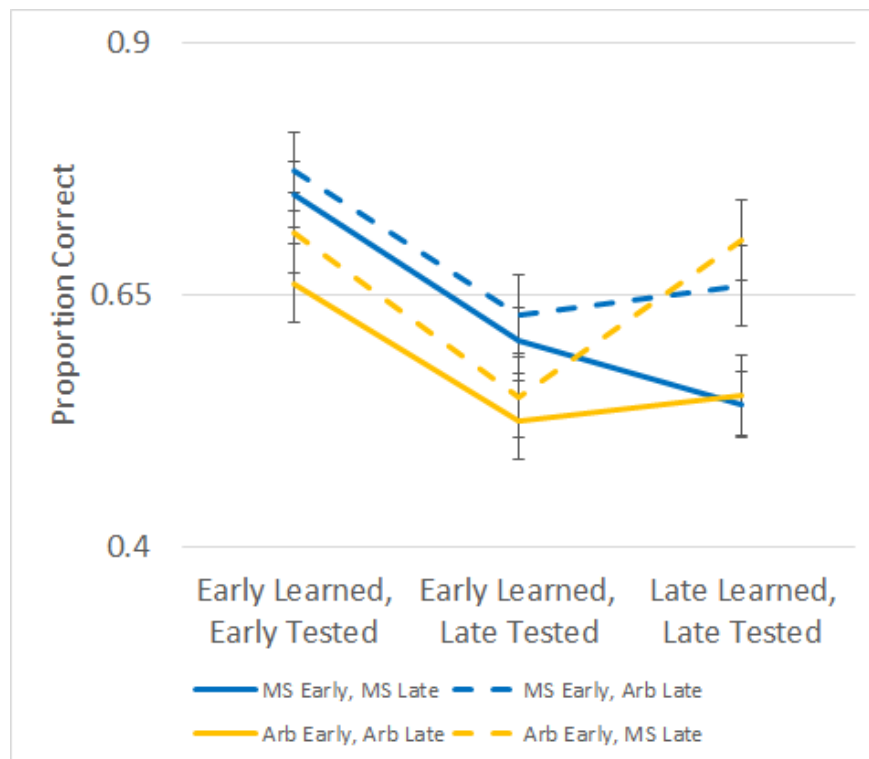


Figure 5.05- Performance of participants in the four conditions of Experiment 1 as a function of the type of word that they were being tested on. The results suggest an overall performance deficit in later trials, reflecting the fact that learning additional words imposes increasing cognitive demands. However, the languages where learning changes (both from MS to Arbitrary and vice-versa) perform significantly better on late learned, late tested trials. Error bars show standard error.

Model comparison revealed a significant effect of experimental condition ($\chi^2(9)=23.44$, $p=0.005$): dummy coding of experimental condition. There was also a significant main effect of Trial Type ($\chi^2(8)=82.75$, $p<0.001$): participants performed significantly better on Early Learned Early Tested trials ($M=0.727$, $SD=0.113$) than on either Early Learned on Late Tested trials ($M=0.58$, $SD=0.085$; $p<0.001$) or

Late Learned – Late Tested trials ($M=0.606$, $SD= 0.107$; $p=0.007$). Finally, there was a significant interaction of condition and Trial Type ($\chi^2(6)= 20.79$, $p=0.002$).

Logistic Mixed Effects Regression II

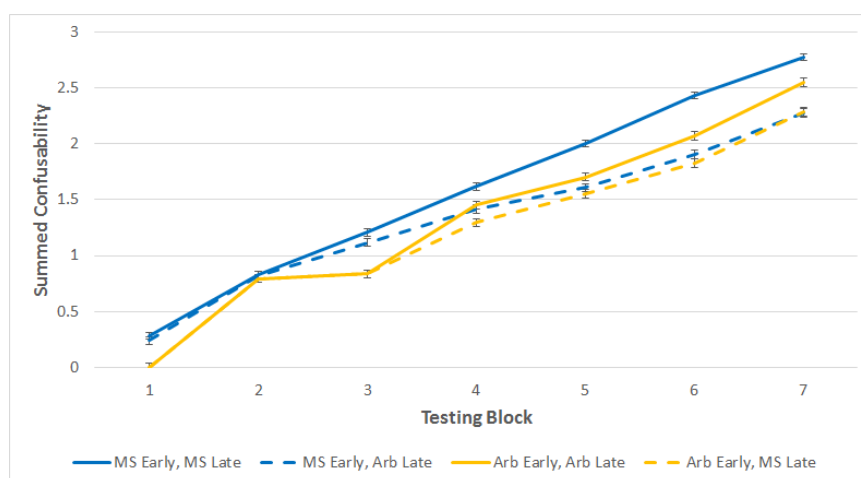


Figure 5.06- Summed confusability metric in each condition across all 7 experimental blocks. Confusability here is a metric to predict performance of participants. Motivated systematic early languages begin more confusable than arbitrary early languages. However, at testing block 4 when the late-acquired language begins being learned and tested, we see that both languages that switch (MS Early->Arb Late, Arb Early->MS Late) have the lowest summed confusability by block 7, predicting that learners of those two languages will perform better than learners of either fully motivated systematic or fully arbitrary languages. Error bars show standard error.

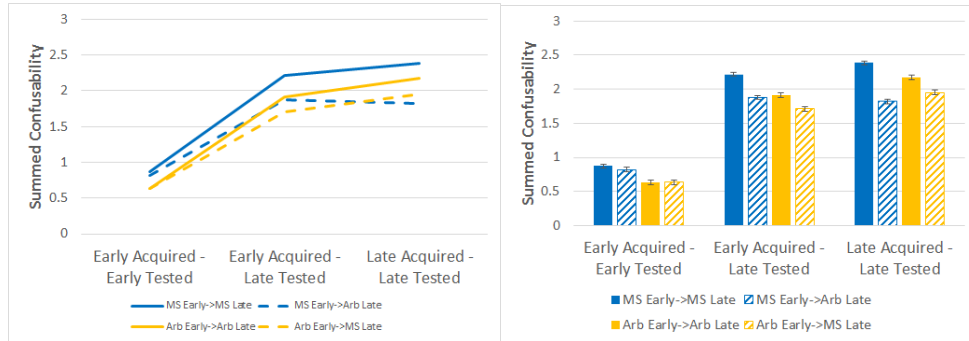


Figure 5.07- A measure of the summed confusability in each condition and trial type. Motivated early language learners have higher summed confusability values, but in late tested trials purely motivated and purely arbitrary languages have the highest summed confusability. Error bars show standard error.

The inclusion of confusability as a factor in my second model eliminated the overall effect of condition ($\chi^2(18) = 24.97$, $p = 0.126$), although there was still a main effect of trial type ($\chi^2(16) = 36.61$, $p = 0.0024$). There was also a significant effect of confusability ($\chi^2(12) = 27.13$, $p = 0.007$), suggesting that my confusability was in fact a good predictor of actual confusability (Figure 5.08).

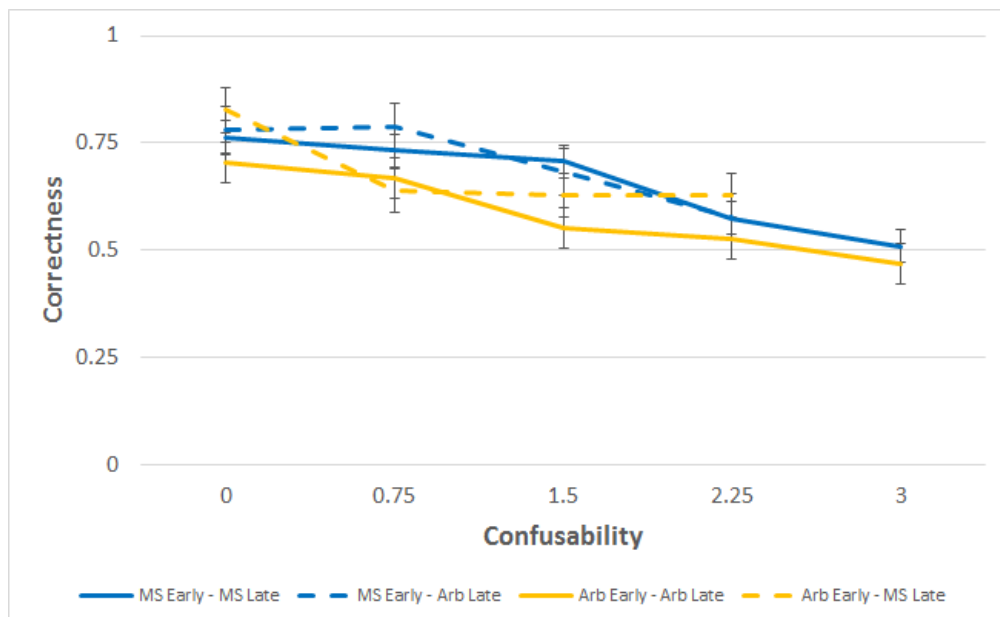


Figure 5.08- Performance results in each experimental condition including confusability as a factor in the model. The results demonstrate that confusability has a main effect: performance is worse on trials where the summed confusability is higher, and no main effect of experimental condition. Missing values on the graph are cells with less than 50 observations. Error bars show standard error.

There were no significant two-way interactions, and only a marginal 3-way interaction of Condition * Confusability * Trial Type significant (all $p > 0.052$).

Test of Bootstrapping

A two sample t test of performance on late-acquired late-tested arbitrary words showed that participants who had learned an initially systematic language ($M = 0.66$, $SE = 0.0415$) performed significantly better than participants who had learned an initially arbitrary language ($M = 0.55$, $SE = 0.0402$; $t = 2.35$, $p < 0.019$).

A second two-sample t-test showed that performance on late-acquired late-tested systematic words was higher for participants who had learned an initially arbitrary language ($M=0.70$, $SE=0.031$) than for participants who had learned an initially systematic language ($M=0.59$, $SE=0.33$; $t= 2.51$, $p=0.013$).

Test of Motivated Early Advantage

Looking at the effect of early language type (motivated systematic vs. arbitrary) on performance on early trials, I found only a marginal effect of early language type: $\chi^2(1)= 3.40$, $p=0.065$ (Motivated systematic $M= 0.759$, Arbitrary $M= 0.687$).

5.3.3 Discussion

Sound Symbolic Bootstrapping

The results of my experiment, at first glance, look to support some version of the simple bootstrapping hypothesis, which suggests that learning motivated associations between words and meanings early increases the subsequent learnability of arbitrary word-meaning mappings. The first test of this prediction comes from comparing late learned, late tested (arbitrary) trials for participants in the MS Early -> Arb Late condition to late learned, late tested (arbitrary) trials for participants in the Arb Early-> Arb Late condition. That comparison seems to suggest that late-learned arbitrary word-meaning mappings are easier for learners who have previous

learned a systematic language than for learners who have learned arbitrary word-meaning associations in the early acquired lexicon, which I found to be true.

However, a closer look at the data suggests that bootstrapping does not actually account for this effect, as its inverse is also true: that is, learners of Arb Early -> MS Late languages perform better on the late learned (motivated systematic) portion of their lexicon than do participants in the MS Early -> MS Late condition.

If we were willing to accept the simple bootstrapping hypothesis based on the data from this experiment, then we would also be required to, paradoxically, accept the possibility that the learning of early arbitrary tokens bootstraps the acquisition of later motivated systematic tokens. Although these two findings are not mutually exclusive (i.e. it is possible that both bootstrapping effects are real), the fact that the data clearly demonstrates both effects suggests the possibility that they might be underpinned by some other variable. Writ broadly, the combination of these two findings suggests that conditions where the early and late learned parts of the lexicon are different result in increased learnability of the later learned part of the lexicon, compared to conditions where early and late acquired portions of the artificial lexicon are entirely systematic or entirely arbitrary. One potential explanation for this finding then is that increasing signal space saturation accounts for the learnability penalty for late acquired late tested word-meaning mappings.

Contrastiveness and Language Learning

The suggestion that more contrastive word-meaning mappings should be easier to learn resulted in two predictions. First, we found, in support of my predictions, that performance varied as a function of the contrastiveness of word-meaning mappings. We can see support for this prediction most clearly in the interaction between Trial Type and Experimental condition: early learned words tested early (when the language is maximally contrastive) had the highest performance, regardless of experimental condition. Additionally, individual word-meaning mappings become more difficult to remember as a function of this interaction: although we see an overall decline in task performance, which can be chalked up to the baseline decline in performance due to general memory constraints, the decline in task performance is less severe for learners of languages that switch from one type of form-meaning mapping to the other (i.e. MS -> Arb, or Arb -> MS; as described above), i.e. those languages where the later learned portion of the lexicon is more contrastive relative to the early learned portions of the lexicon.

The results of the second linear mixed effects model, which included confusability as a predictor, eliminated the main effect of condition, making it clear that confusability accounts for the decrease in learnability across experimental conditions. Regardless of experimental condition, performance was significantly worse on less contrastive labels.

The second analysis including confusability as a predictor also included a significant main effect of trial type, with participants performing better on early learned – early tested trials than the other two trial types. Several explanations seem immediately plausible for this finding. First, the benefit for early learned – early tested trials might reflect a primacy effect that is eventually washed out by increasing confusability. Second, the benefit might be due to early learned – early tested trials in motivated systematic early languages providing a benefit for their learners: however, we found no significant interaction between Trial Type and Experimental condition, and a post-hoc test looking at motivated systematic vs. arbitrary early trials did not suggest an effect of motivatedness. Finally, early learned-early tested trials have fewer labels to learn, and a smaller test array of possible choices.

5.4 General Discussion

The results presented here support the general conclusion of the experiments presented earlier in this dissertation and the findings of previous researchers with regards to the benefits and costs of systematicity: systematic associations between words and meanings enhanced the learnability of those tokens, but the degree to which this was true varied as a function of the overall signal space saturation and/or relative contrastiveness of each word to other words of its type. Early acquired systematic associations between words and meanings were learned most easily,

although the degree to which this learnability benefit might have been produced by motivatedness, rather than systematicity is impossible to determine given the data for this experiment, although the results of the experiments from Chapters 2-4 suggest that both are candidates for explaining the increased performance. We also demonstrated, for the first time in an artificial language learning experiment of this type, that potential confusability penalties for systematic word-meaning mapping vary not only as a function of the contrastiveness of entire artificial lexica (as in Monaghan et al., 2011; and Experiments 1-3 from Chapter 2), but also that the introduction of new labels that are similar to existing labels can lead to confusion even on previously well-learned word-meaning pairs.

The findings of this experiment are particularly relevant for proponents of bootstrapping hypotheses in general, as they demonstrate the possibility that invocations of bootstrapping hypotheses might suffer from the *post hoc, ergo propter hoc* logical fallacy. Assuming that there is some selection process operating over languages such that word-meaning mappings are chosen in an optimized fashion, the pressures of contrastiveness might necessitate the shift from an early motivated and/or systematic lexicon to a less constrained arbitrary lexicon, but the fact that these two stages of language learning occur in succession would not necessarily suggest that the first influences the learnability of the second in any way that could be described as “bootstrapping” or “scaffolding” that learning process. Assuming that it is true that in natural languages earlier acquired parts of the lexicon are

systematic (and potentially motivated) whereas the later acquired parts of the lexicon are more arbitrary does not suggest causality- given the interaction of two selection pressures: contrastiveness maximization and learnability enhancement (due to systematic structure, motivatedness, or both) the fact that arbitrary and systematic associations between words and meanings are favored at different times does not suggest that one causes the other. *Referential* and *conceptual* bootstrapping are, despite seeming more tenable than *simple bootstrapping* hypothesis explored here, not inured to this possibility: it is difficult to determine the degree to which later learning is contingent on, and thus bootstrapped by earlier learning. The use of motivated word-meaning mappings might, for example, ease the ability of naïve language learners to establish reference between sound and meaning, but the degree to which this established reference is actually generalizable to non-motivated tokens is much more difficult to establish.

5.4.1 Extensions

The experimental protocol here offers a number of opportunities for testing many of the hypotheses raised in the artificial language learning and sound-symbolism literature generally. The inclusion of non-motivated systematic languages, for example, might help further enrich the results of the experiments presented in Chapter 4 – for example, we do not currently know if motivated mappings continue to be beneficial beyond the earliest exposure to tokens.

Similarly, although the results of this experiment give better insight into the influence of contrastiveness on learnability, the ability to look at the transition between levels of contrastiveness that favors systematicity or arbitrariness was limited by a number of factors. The method of construction for arbitrary word-meaning mappings in this experiment was still relatively constrained – although previous experiments, e.g. Monaghan et al. (2011) used arbitrary signal spaces that were even more tightly constrained (and more similar to the systematic signal spaces) it is possible to work with arbitrary associations that are significantly more contrastive than the one used here. In a maximally contrastive arbitrary language, the learning of new labels should not interfere with performance on older labels, other than an impairment due to increasing cognitive demand and task difficulty- here, however, even my arbitrary word-meaning mappings were relatively constrained and became increasingly confusable over time (though not as quickly as the systematic labels. Using a larger signal space and more contrastive labels for the creation of arbitrary languages would allow for an extension of this experiment, because an Early Arbitrary -> Late Arbitrary language could serve as a baseline to which other conditions could be compared, establishing not only the degree to which increased task demands influence learnability, but also e.g. the influence of primacy and recency effects.

5.5 Conclusions

In general, we find further support for the notion that systematic associations between words and meanings can benefit language learners under certain conditions, i.e. those where the benefits of systematicity accrued due to the similarity of words for related meanings are not overwhelmed by the potentially confusability induced by that same similarity. The interesting insight from this experiment in particular is that word-meaning pairs that are already established in the lexicon of their learners can be interfered with by the introduction of additional labels, and the degree to which learning new labels penalizes the memorability of previously learned labels varies as a function of the similarity of those labels. Previous to these findings one could imagine, for example, that the individuation penalty for less contrastive systematic languages would only be incurred by newly learned words, and that this penalty would be sufficient to push languages towards arbitrary word-meaning mappings so that those new words could be learned more easily. Our results, however, suggest that in addition to a pressure for new words to be more contrastive for the benefit of their own learnability, existing words also suffer from penalties to memorability under less contrastive conditions – this suggests that in addition to a pressure for new labels to be learnable, languages suffer a secondary pressure towards arbitrariness, in that already learned words can be negatively impacted.

The results of this experiment also highlight the fact that bootstrapping explanations, including those offered in this dissertation, must meet a difficult burden of proof, lest we risk committing a *post hoc, ergo propter hoc* fallacy. Selective pressures on the expressivity, usability, or transmissibility of language might account for the fact that early acquired portions of the lexicon are more likely to be systematic or motivated and also for the arbitrariness of subsequently learned words, without any enhancement of later learning that can be rightfully described as bootstrapping. The finding that confusability predicts learnability and accounts for some of the differences in learnability between systematic and arbitrary lexicons also suggests a number of interesting but results in the existing literature in the field, but also to explore non-arbitrary configurations of the language that might have the benefits of systematicity without inducing further confusability (i.e. while maintaining contrastiveness).

Chapter 6

Conclusions

In this dissertation I have undertaken an exploration of motivatedness and systematicity, and the effects that those type of non-arbitrary associations between words and meanings have on learning and the structure of the lexicon. Above the level of the lexicon, language is recognized to be shaped by pressures to make it more learnable, expressive, and communicatively functional (Kirby et al., 2015). The recognition that these pressures might influence the structure of language are relatively new, especially when stated explicitly, but all attempts to delineate universals of human language (e.g. Hockett, 1960) are fundamentally related to these issues. By exploring features common to all languages, previous researchers have necessarily found themselves describing the outcomes of those pressures. Further, because what is learnable, expressive, or communicatively functional is determined by the perceptual and cognitive organization of language learners, the exploration of these language universals also tells us important facts about human cognition more generally.

The acknowledgement that cognitive biases, especially those that are domain general, have downstream effects on language has proven to be a rich source of

explanatory power for the structure of human languages above the level of the lexicon. More systematic languages are more compressible, and thus easier to learn (Chater & Vityani, 2003; Tamariz & Kirby, 2015), and the structure of syntax reflects the structure of events (Haiman, 1980, 1985). Human language learners share the same basic perceptual and cognitive structures, and thus human languages, despite being different, share some features that reflect the strengths and constraints of human perception and cognition (Hockett, 1960). The pressures for language to be expressive but learnable, which are mediated by these shared perceptuocognitive features, likely accounts for the shared features of many languages. It is, however, important to recognize that not all languages produce the same solution to the pressures for expressivity and learnability.

As we have seen in both the psycholinguistic literature broadly (e.g. Monaghan et al., 2011; Perniss & Vigliocco, 2014; Dingemanse et al., 2015) and in this dissertation, the pressure for languages to be learnable can, under some circumstances, favor arbitrariness and explain the predominance of arbitrary word-meaning associations. Hockett (1960) recognized both arbitrariness and learnability as universal features of human language, but we might instead suggest that learnability is a language universal on its own, but also one that accounts for the form that other language universals take, especially when we also consider the pressure for expressivity. Languages have universal features like discreteness because those features make languages more learnable and/or expressive:

learnability is both a design feature and a pressure that accounts for the presence of other design features. Human perceptual and cognitive biases mediate the pressures of learnability and expressivity, and thus shape the expression of language universals, and this recognition has been productive for exploring questions about the evolution of language.

At the level of the lexicon we can see a number of similar perceptual and cognitive biases. Motivated associations between words and meanings arise because of perceptual biases, while domain-general memory and other cognitive constraints bias towards systematicity and compressibility. The presence of non-arbitrary word-meaning associations has been recognized for a long time, but often treated as marginal (Newman, 1933, Newmeyer, 1993; Saussure, 1983). In recent years, however, researchers have increasingly suggested that these associations are probably important (Nielsen & Rendall, 2012): they might make words more learnable (Nygaard et al., 2009), expressive (Yardy, 2010) or both (Nielsen, 2011). Here, I have explicitly suggested that the way that perceptuo-cognitive biases influence the structure of the lexicon is homologous to the way that those same biases influence syntax and morphology: perceptual and cognitive biases determine the features that favor learnability, and those features, through an interaction with the pressure for expressivity, determine the structure of the lexicon.

In the remainder of this chapter I will briefly review the model for the contribution of motivatedness and systematicity to the structure of the lexicon that was presented in Chapter 1, focusing on the empirical evidence presented here and elsewhere in support of that proposition. Ultimately, this model is unlikely to explain everything about the structure of the lexicon, but modestly it can serve as a platform to motivate future research. Where the model currently falters in its explanation of structural outcomes or their timing, I will point towards some plausible directions for future research.

6.1 Motivatedness, systematicity, and language learning

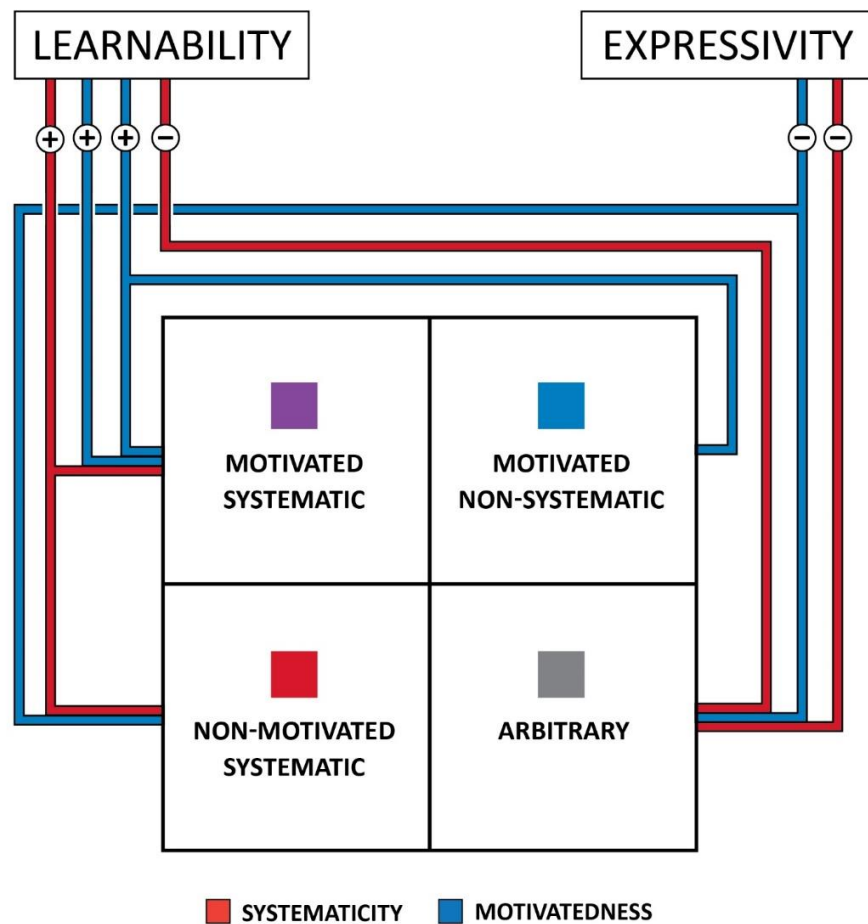


Figure 6.01- A robust model of the pressures for learnability and expressivity and their contribution to the lexicon.

To rehearse briefly, previous research has suggested the non-arbitrary associations between words and meanings might have important implications for the learnability of languages (e.g. Monaghan et al., 2011). Motivated associations between words and

meaning have been suggested to allow for reference to be established (Imai & Kita, 2014): here I have called that proposal the *referential bootstrapping hypothesis*. This hypothesis suggests that the use of motivated word-meaning mappings to establish reference can be generalized to enhance the establishment of reference in non-motivated cases. Support for this hypothesis comes from a number of sources, although it has not been tested directly experimentally. First, children are able to learn motivated associations between words and meanings more easily than arbitrary associations (Asano et al., 2015), and this is true cross-linguistically (Kantarzis et al., 2011). Additionally, adults have been shown to demonstrate the same effect in artificial language learning paradigms (e.g. Nygaard et al., 2009). In languages that make use of large classes of non-arbitrary words like ideophones, motivated word-meaning mappings are learned earlier and more easily than their arbitrary counterparts (Imai & Kita, 2014), suggesting that perceptual bias is being leveraged to enhance learning.

Research exploring motivated word-meaning mappings has exploded in the last 5 years, and invocation of ideas like *referential bootstrapping* to explain the benefit that motivated mappings might have for language learning more generally has become increasingly common (Imai & Kita, 2014; Perniss & Vigliocco, 2014; Dingemanse et al., 2015). Here, however, I offer one additional suggestion: that motivated incidentally systematic mappings between sets of words and sets of meanings might make concepts and categories underlying the structure of the

lexicon more obvious: a proposal that I have called *conceptual bootstrapping*. The fact that much of the artificial language learning literature exploring the learnability of motivated word-meaning mappings like the Bouba-Kiki effect are also structured such that they are incidentally systematic (e.g. Aveyard, 2012; Monaghan et al., 2012), however, makes determining the effect of motivatedness on learning slightly more problematic.

Research exploring the effect that systematicity at the level of the lexicon has on learning, and the degree to which natural lexica are systematic, has been much less common than research into motivatedness, but has resulted in similar suggestions: systematic associations between words and meanings might increase learnability, but the degree to which this is true is likely to depend on the nature of the systematic associations and how greatly those associations impinge on the language's contrastiveness (which captures both expressivity and learnability)(Monaghan et al., 2011).

Thus, previous research left a number of questions unanswered that the research presented in this dissertation attempted to address:

- 1) What effect does systematicity have on contrastiveness, and thus on learnability?
- 2) How do different realizations of systematic word-meaning mappings influence learnability?
- 3) What are the benefits of motivatedness for language learning? Are these real effects, or are they mediated by (incidental) systematicity, rather than motivatedness?

- 4) What is the explanatory value of bootstrapping hypotheses? Is there evidence for bootstrapping, or do other features like contrastiveness best explain previous findings?

6.2 Summary

6.2.1 Experimental Evidence

Chapter 2

In Chapter 2, I presented the results of a series of experiment designed to examine, as straightforwardly as possible, the effect of systematicity on language learning. I found, following previous researchers, that systematic associations between words and meanings provide a benefit for language learning for categorization, but can penalize individuation learning. However, the use of two different sets of pseudowords between Experiments 1 and 2 allowed me to demonstrate that the degree to which systematicity can actually penalize learning varies as a function of the confusability of pseudowords of the same type.

In Experiment 1, which used a set of pseudowords constructed similarly to those in Monaghan et al. (2011) we found that learners of systematic languages performed well on tasks that were aided by having a transparent category structure. On out-of-class distractor trials, systematic language learners were able to quickly and easily reject pairs of words and meanings that were not coherent with the category structure of their language (e.g. rejecting ‘mo nu mu’ as the label for a

vehicle). The performance of those same systematic language learners on in-class distractor trials, however, suggested that they had failed to learn the names for any individual meanings. When presented with a word that was of the correct type, but not the actual correct word (e.g. ‘mo nu mu’ for an animal, but not a badger) systematic language learners performed significantly below chance: they had learned the category structure of the language, but had somehow failed to learn individual words. This suggested that something about the systematic mapping between words and meanings interfered with the process of individuation.

In Experiment 2, where the chosen languages were maximally contrastive, systematic language learners did not have this problem: they were still able to reject out-of-class distractors at similar rates, but did so while maintaining the ability to learn individual words. This result suggested that the contrastiveness of words to one another influenced learnability: when languages could be systematically marked in such a way that individual words within systematic categories were still distinct, the language could aid performance on out-of-class distractor trials without a commensurate decrease in the ability to individuate (Figure 6.02).



Figure 6.02- Results from Experiments 1 and 2 demonstrate that the benefit accrued to systematic language learners is both contingent on contrastiveness. When words are less similar to one another, they are less easily confused and thus easier to learn. Error bars show standard error.

In Experiment 3 of Chapter 2, I attempted to replicate the findings of Monaghan et al. (2011)'s 3rd experiment, introducing half-half languages that were systematically marked, but less constrained. I suggested that Monaghan et al. (2011)'s results were interesting, but underpinned by contrastiveness in a way that was not straightforwardly captured by their experimental design: i.e. the half-half languages that they created were indeed more learnable because they were more contrastive, not necessarily because of their partially systematic construction. First, these

languages were still, depending on how they were analysed, fully systematic: the same phonemes did not appear in the same locations between the two types of languages. Second, the chosen words for the half-half language were actually maximally contrastive: the manipulation of word construction into partially systematic types allows for the possibility to select more contrastive languages, but does not ensure it. I attempted to address these criticism by introducing an additional factor of signal space size, suggesting that lexica chosen from a smaller signal space would be inherently less contrastive and thus more difficult to learn. I did not, however, find support for this proposal: the manipulation of signal space size failed to capture contrastiveness in a way that was stable enough to allow for exploration. Specifically, the results of this third experiment highlighted the difference between signal space size and saturation: given an equal number of tokens larger signal spaces will on average be more contrastive, but this is not guaranteed. Experiment 3 was also designed to include a contrastiveness measure for each pseudoword that I hoped would predict performance, but ultimately failed to capture the feature of interest (In Chapter 5, I returned to an exploration of contrastiveness using a new metric).

Chapter 3

In Chapter 3, I explored the proposal that decreasing contrastiveness negatively impacts learnability more indirectly. The systematic associations explored in

Chapter 2, as well as those explored by Monaghan et al. (2011), rely on systematically mapping features of phonology (e.g. plosiveness) to features of meaning (e.g. animals) such that similarity on the signal dimension maps onto similarity on the meaning dimension. In Chapter 3, I explored the degree to which experimental results could actually be traced to phonological features. I suggested, following the results of Chapter 2, that a complete explanation of how contrastiveness affects learnability would require a better understanding of what features are salient to human language learners. Thus, in Chapter 3 I introduced a phonological clustering factor to explore systematic word-meaning mappings that were not based on phonological features, but that were still systematic when analysed based on their phoneme inventories. Thus, instead of matching all plosive words with animals and all sonorant words with vehicles, I created pseudoword categories that were still systematic, but where phonological features were not predictive. I explored the effects of this manipulation both using the same experimental methodology as in experiments 1-3 and using a replication of Monaghan et al. (2011)'s model.

Our replication of Monaghan et al. (2011)'s model suggested that phonologically dispersed languages, where dissimilar phonemes were mapped onto similar meanings, were easier to learn: maximizing categorisation performance while minimizing the individuation penalty in much the same way that I found in Experiment 2 from Chapter 2. Following Monaghan et al. (2011), I used the results

of our model to make predictions about the performance of my human language learners. I found, however, counter to the results of the model, that a contrastiveness metric based on phonological features was not predictive of the performance of human participants (Figure 6.03).

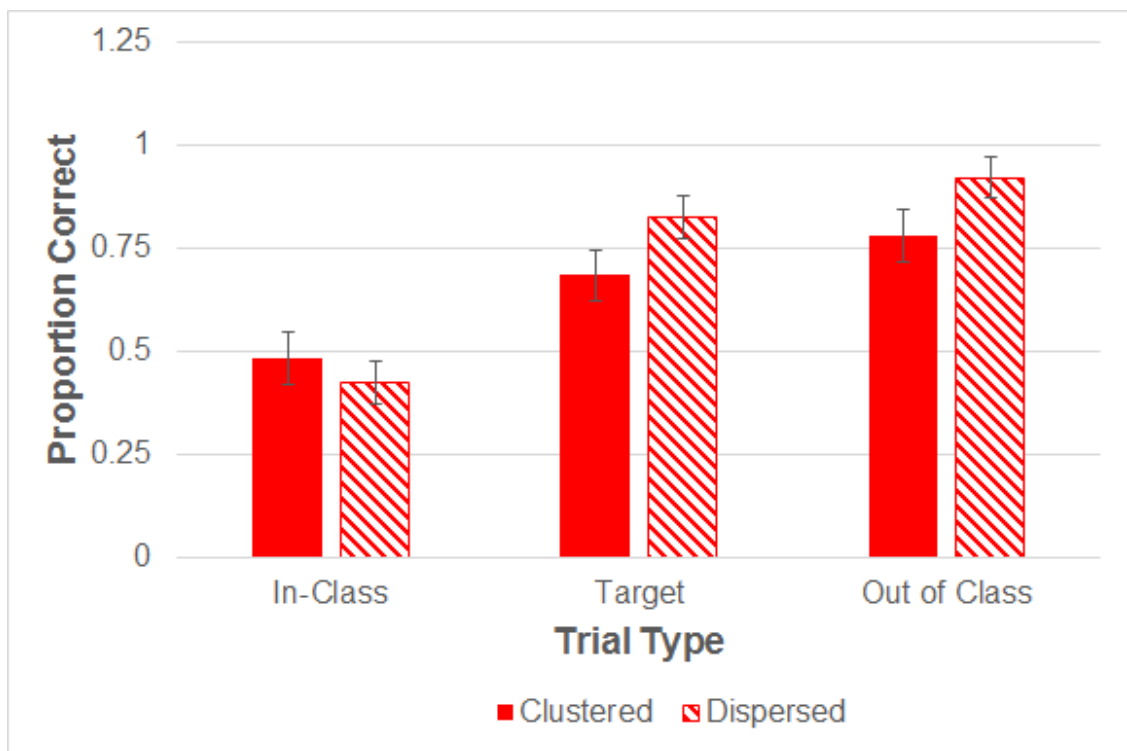


Figure 6.03- Results from Experiments 4 (Chapter 3). For human participants there was no effect of phonological clustering, suggesting that phonological features were not predictive of learnability. Error bars show standard error.

The difference in performance between our model and human participants can be traced directly to contrastiveness: specifically to the determination of what features are relevant in determining similarity, and thus confusability, for the model and the

human learners. The phonological feature representation used by the model ensures that similar phonemes are represented very similarly to one another, /g/ and /k/ for example differ only in voicing in the model, and thus are very similar to one another and easily confused. For human learners, /g/ and /k/ still differ only in terms of voicing, but are nonetheless recognized as separate phonemes whose difference is highly relevant for the language. The results of Chapter 3 thus suggest that: a) systematic associations between words and meanings do not need to be based on phonological similarity to influence learnability; b) human language learning is, in this context, better explained as being mediated by similarity based on phonemes being discrete, rather than by assuming that phonemes are clusters of phonological features; and, c) that conclusions drawn from computational models should be considered carefully.

Chapter 4

In Chapters 2 and 3, in addition to an exploration of systematicity generally, I suggested that the *categorisation* metrics used in Monaghan et al. (2011), and thus in our own model, were not actually appropriate metrics of categorisation, because they reflected errors in response to a task that was explicitly about individuation. Systematic associations between words and meanings might increase the ability of human learners to categorise, but neither Monaghan et al. (2011) nor the experiments in Chapters 2 and 3 actually tasked experimental participants with

categorizing. In Chapter 4, we presented the results of a task that is explicitly about categorisation, rather than individuation.

To explore categorisation, I departed from a comparison of the learnability between systematic and arbitrary languages to compare the ability of learners to categorise correctly with languages that were systematic, but either motivated or non-motivated. Previous research (Aveyard, 2012; Monaghan et al., 2012; Nielsen & Rendall, 2012) has suggested that motivated associations between words and meanings are more easily learned than are their counter-motivated counterparts. These results, however, are difficult to analyse with respect to motivatedness enhancing learning over arbitrariness. First, the motivated associations used in these experiments were also incidentally systematic, and second their learnability was compared to counter-motivated, rather than arbitrary tokens. To this end, we conducted two experiments exploring the difference in learnability between motivated (incidentally) systematic and non-motivated systematic languages and determined that there is a learning benefit for motivatedness, and that that benefit comes in the earliest testing trials where naïve expectations based on perceptuocognitive biases allow participants to answer correctly prior to any learning (Figure 6.04).

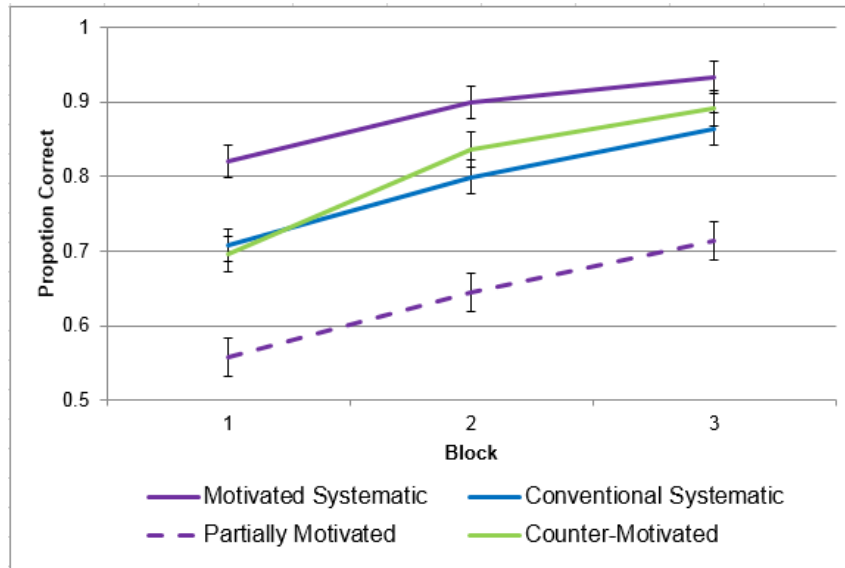


Figure 6.04- Results from Experiment 5 show that motivated systematic languages are easier to categorise than conventional systematic languages, but only on early trials. Error bars show standard error

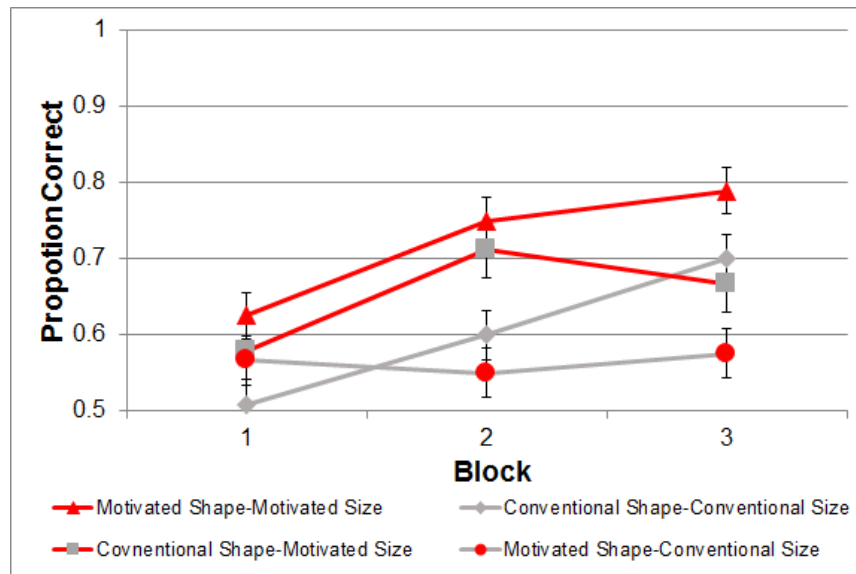


Figure 6.05- Results from Experiment 6 show that the presence of a motivated association between features of the word and features of the meaning on one dimension negative influences the learnability of a non-motivated association for a second, unrelated feature of meaning. Error bars show standard error.

Additionally, we found that the presence of motivated associations between words and meanings interfered with the learnability of arbitrary associations on a second, unrelated dimension (Figure 6.05). We suggested that this finding might help account for the relative lack of motivatedness in natural lexica, or at least help explain why non-arbitrary parts of the lexicon are isolated from the rest of the lexicon in some languages like Japanese (Asano et al., 2015). Ideophones and expressives (Akita, 2011), for example, are noted for their markedness (Newman, 2001), being described as being phonologically aberrant or peculiar (Newman, 1968; Epps, 2005; Kruspe, 2004) or structurally marked (e.g. Klammer, 1999). The markedness of these ideophones, which effectively insulated them from the rest of the lexicon, might exist to stop the presence of these associations from negatively influencing the learnability of arbitrary words.

Chapter 5

In Chapter 5 we introduced a temporal component to the artificial language learning paradigms used in previous chapters to allow for an exploration of the way that learning changes over time. Specifically, the introduction of this temporal component allowed me to evaluate the possibility of naïve bootstrapping: i.e. the suggestion that learning non-arbitrary word-meaning mapping bootstraps learning of later-acquired arbitrary words (but not based on either referential or conceptual

bootstrapping). I found evidence that appeared to initially support this possibility, but that could actually be more properly traced to confusability: systematic languages, as they grow, become increasingly confusable, and this eventually swamps the learning benefit that they gain from being systematic in the first place (Figure 6.06).

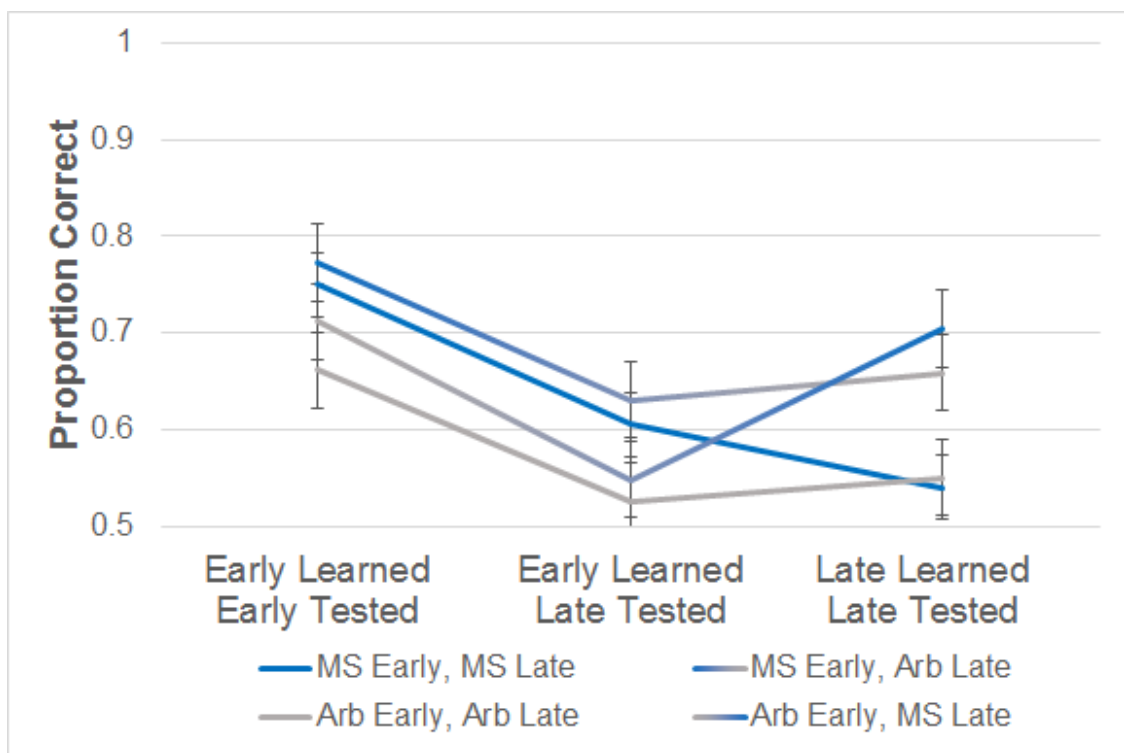


Figure 6.06- Results from Experiment 7 show that systematic languages become increasingly confusable as the number of words to be learned increases. Error bars show standard error.

In my growing lexicon experiment, I found that later-learned words were learned more poorly overall because the task of learning more words is inherently more difficult, but that learning words that were more contrastive later increased their

learnability. In the case of learners who moved from learning motivated systematic words to arbitrary ones, this appeared to support a bootstrapping hypothesis, but the same results were found for participants moving from arbitrary early lexica to later motivated systematic ones (Figure 6.07).

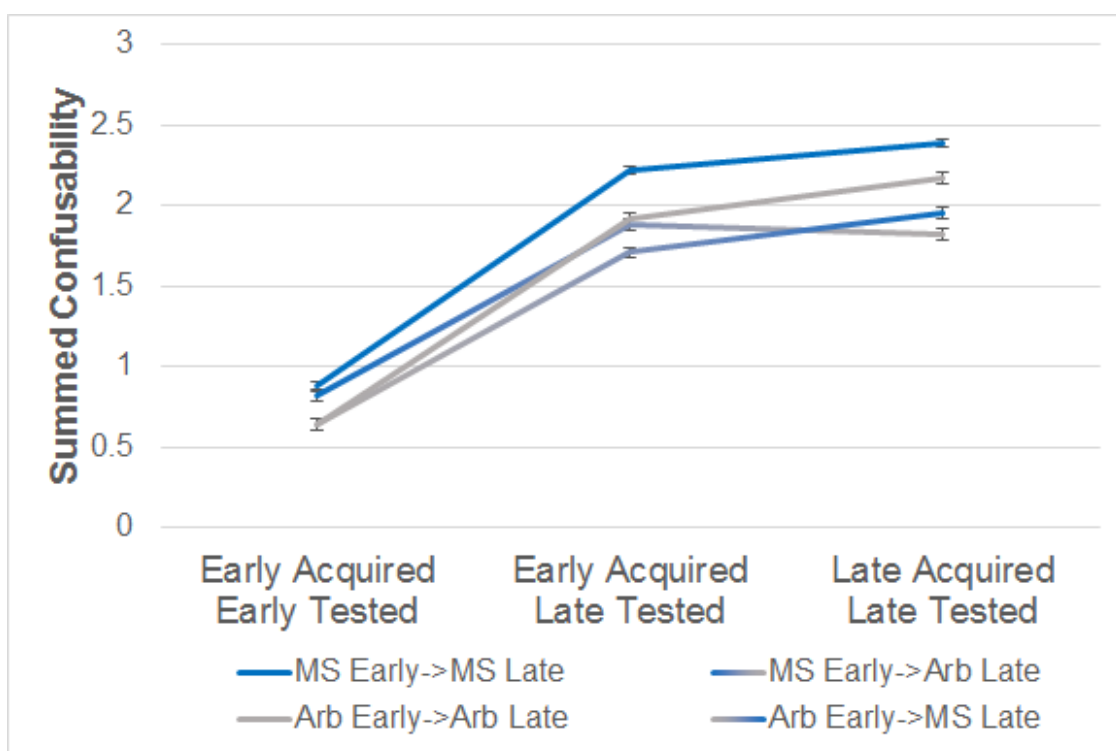


Figure 6.07- The summed confusability metric from Experiment 7 shows that languages that contrastiveness, rather than bootstrapping, predicts the difference in learnability between the four language types from that experiment. Error bars show standard error.

6.2.2 Bootstrapping

My overall results, especially those of the growing lexicon experiment presented in Chapter 5, suggest that I should be critical of bootstrapping proposals for the benefit of non-arbitrary word-meaning mappings for language learning. Although I was unable, using adult participants, to test for the possibility of either referential or conceptual bootstrapping, my results still suggest that the temporal trajectory of natural language learning does not necessarily suggest bootstrapping or scaffolding. Both motivatedness and systematicity have clear benefits for learning under certain conditions, and it appears that these conditions are best met in early language learning: new learners can use motivated associations to establish reference, and systematic mappings allow for generalizability that benefits some types of learning. The existing evidence suggests that the structure of natural lexica reflects these benefits: early acquired portions of the lexicon are indeed more systematic (Monaghan et al., 2014) and motivated (Asano et al., 2015) than the later acquired arbitrary remainder of the lexicon.

As I stressed in Chapter 5 however, the mere fact that acquiring arbitrary word-meaning pairs occurs after the early acquisition of more non-arbitrary words does not imply that the learning of the first enhances the learning of the second. Even the referential and conceptual bootstrapping hypotheses, which I was unable to test directly, rely on the suggestion that the learning of the non-arbitrary words

enhances the learning of later arbitrary ones. However, the fact that more difficult to learn arbitrary associations are learned later could instead reflect general cognitive development reaching maturity and then being brought to bear on the more difficult learning task, rather than early-acquired words accounting for the enhancement. So, we must recognize that bootstrapping hypotheses are susceptible to *post hoc ergo propter hoc* reasoning, and this is especially true when we consider contrastiveness as a pressure that can significantly shape learning trajectories.

6.2.3 Contrastiveness and Confusability

Collectively, the results of the experiments presented above suggest that, much like language above the level of the lexicon, the pressure for languages to be learnable accounts for the general structure of the lexicon. The conditions under which a language is learnable are determined by the perceptual and cognitive organization of its learners: certain types of associations are more learnable by virtue of their being perceptually biased, and constraints from domain-general systems like memory similarly influence what kinds of associations and structures of associations can be learned. Here, we have suggested, following others (Gasser, 2004; Monaghan et al., 2011), that both motivatedness and systematicity can enhance learning, but that the constraints that these non-arbitrary mappings impose on the available signal space create the conditions that limit the degree to which they can be beneficial for learning.

Motivated associations between words and meanings do not individually constrain the signal space to any large degree, although a language based only on motivated associations would suffer from limited expressivity because it would only be able to express a limited number of concepts. However, as I have suggested, motivated associations can become incidentally systematic, and this systematicity constrains the size of the signal space.

In the case of systematicity, mapping similar words to similar meanings can benefit learning, especially in cases where categorization is relevant: at the very least systematic associations limit the cognitive load required to discount out-of-class pairings. At the same time however, increasing the number of words in a given signal space increases the possibility, given some error, that words will be confused for one another. This increase in confusability as a function of signal spaces becoming increasingly saturated accounts for the majority of my findings, and suggests that languages will favor systematicity only insofar as systematic word-meaning mappings do not result in confusable word-meaning pairings (over and above some baseline level of confusability based on simply learning more words).

6.3 Future Directions

6.3.1 Bootstrapping

The findings of the experiments presented in this dissertation, as well as the more general claims outlined here, point towards a number of potentially profitable directions for future research.

First, although I am critical of invoking bootstrapping explanations based simply on temporal order of events, both referential and conceptual bootstrapping are plausible and account for the observed data in human language learners fairly well. Unfortunately, because adult learners have already learned to establish reference, exploring the referential bootstrapping hypothesis using typical experimental participants and methodologies might be impossible. However, Imai & Kita (2014)'s *sound symbolism bootstrapping hypothesis* is ripe for empirical testing: we already have evidence that infants attend to motivated word-meaning associations, and require only observations demonstrating that attention to motivated associations can be leveraged to establish reference for arbitrary ones.

The conceptual bootstrapping hypothesis suggests that learning non-arbitrary associations between words and meanings can enhance the ability of learners to recognize concepts and categories that are relevant to their language and can be leveraged for later language learning. Although adult learners have already

established the ability to recognize categories and generalize across those categories (ref?) they might still benefit from non-arbitrariness in the establishment of new categories. Experimental stimuli like the “Yufo” (Gauthier and Tarr, 1997), where the distinction between the two types of images is not immediately apparent, even to adult learners, might allow for the best test of categorisation. Previously, authors have suggested that labelling superordinate categories generally allows children to learn to form those categories (Waxman & Hall, 1993; Waxman & Markow, 1995), and that relational concepts underpinning these categories can then be transferred to novel stimuli (e.g. Ratterman & Gentner, 1998). With systematicity at the level of the lexicon however, we are not interested in superordinate terms, but rather in how similarity within categories (or motivatedness of association) might similarly influence category formation. The “Yufo” stimuli used in Lupyan, Rakison, & McClelland (2007; Figure 6.08) are well suited to this task because the distinction between the two types is not immediately apparent, even to adult learners.



Fig. 1. The two categories learned by the subjects. The stimuli on the left have flatter bases and a subtle ridge on their “head”; the stimuli on the right have more rounded bases and a smoother head.

Figure 6.08- “Yufo” stimuli from Lupyan et al., 2007.

Lupyan et al. (2007) found that simply by having names, the category distinction between the two types of Yufos was made more salient and the categories were learned more easily, despite the fact that the inclusion of names required additional learning. In 2014, Lupyan & Casasanto returned to these stimuli, demonstrating that when the superordinate names for the two types of yufos were motivated (‘foove’ for round-headed yufos and ‘crelch’ for pointier yufos) categorization became easier. This result certainly seems to suggest that motivatedness, at least, might bootstrap category formation. I propose a simple extension of this experimental paradigm where names are given to these stimuli directly, rather than

labeling only their category. This manipulation would allow for a test of whether motivatedness and systematicity can more generally bootstrap the acquisition of categories, and further whether these learned categories can then be generalized to more arbitrary labels.

The experiment suggested above addresses the claim that non-arbitrary word-meaning mappings might facilitate the learning of category boundaries, but what about the establishment of categories more generally? Again, adult participants are already aware of the fact that the objects in their language can belong to meaningful categories, but what about children? The conceptual bootstrapping hypothesis suggests, in addition to making relevant dimensions more salient for adult learners, that this saliency might underpin the recognition that categories exist at all, much in the same way that motivated word-meaning mappings can be suggested to underpin the establishment of reference. Testing this possibility requires infant participants, but might otherwise use a similar methodology to Lupyan & Casasanto (2014) (although preferential looking, rather than direct responding, would likely be required).

6.3.2 Contrastiveness

Because different languages likely have differently sized signal spaces, the search for a specific optimal configuration of language that would allow for maximal benefits based on systematicity and motivatedness while maintaining sufficient

contrastiveness would be a fool's errand. However, the recognition that systematicity and motivatedness might inherently limit contrastiveness still suggests a number of directions for future research.

An understanding of the relationship between signal space saturation, systematicity, and confusability both explains some of the existing findings in the psycholinguistic literature and points towards predictions about the structure of the lexicon that are so far not attested. First, I suggest that because non-arbitrary mappings between words and meanings limit the available signal space, and because additional systematic dimensions further limit the available space, individual natural languages should leverage non-arbitrariness differently.

Motivated associations between words and meanings may enhance learning, especially in early acquired words, but this does not suggest that the sound-symbolic mappings that we observe in one language should necessarily be found in all (or even most) other languages. Because there are many possible crossmodal associations that can be leveraged linguistically to increase the salience of certain word-meaning mappings we should instead expect that each language will arrive at a similar overall solution for how to leverage motivatedness without penalizing learnability or expressivity, but that the specific motivated dimensions leveraged for this purpose will be somewhat random. The results of Experiment 2 in Chapter 4 of this dissertation point towards why languages might not actually be most learnable if all

possible motivated associations were manifest in the language: the presence of a motivated association on one dimension might actually interfere with the learning of arbitrary associations. Because not all meanings that a language expresses are equally (or at all) amenable to motivated mappings, the presence of too much motivatedness might actually limit learnability.

Similar suggestions might be made for systematic associations at the lexical level: some systematicity is good, but only insofar as a language remains sufficiently expressive and learning is not penalized due to increased confusability. But, different languages will arrive at different solutions for where systematic word-meaning mappings can be best leveraged, although some categories will obviously be more relevant early in language acquisition, and thus more likely to be systematically structured.

These two suggestions are especially important if one takes bootstrapping hypotheses seriously. Under either referential or conceptual bootstrapping, only a limited number of non-arbitrary word-meaning associations would be required to scaffold language learning, and as such the idiosyncratic use of non-arbitrary mappings between languages would seem less strange.

Exploring these possibilities in natural languages seems daunting and potentially tautological: how could the fact that different languages leverage different motivated or systematic mappings suggest that those mappings represent

different solutions to the same overarching challenge of optimization? The use of artificial language learning tasks, especially iterated learning protocols that allow for non-guided optimization seem to be the most profitable way to explore these possibilities. Initial generations of participants might be trained with languages that used a large number of non-arbitrary word-meaning mappings, and have the learnability of this language compared with the output of later generations: how many non-arbitrary mappings might be maintained in later versions of the language, and what dimensions might prove to be most favorable for the persistence of these non-arbitrary mappings?

Finally, considerations of the interaction between contrastiveness and non-arbitrariness might help explain the isolation of ideophones from the rest of the lexicon in languages that have large numbers of non-arbitrary word-meaning mappings. By using a portion of the possible signal space that is not otherwise utilized by the lexicon, these languages might gain all of the benefits of systematicity and motivatedness without materially influencing the contrastiveness of the remainder of the lexicon. In terms of my findings from Chapter 5, this configuration would still result in a penalty for learnability as more non-arbitrary words are learned (early acquired words would be easier to remember when there were few words, but would become more confusable as additional non-arbitrary words were introduced), but the introduction of non-arbitrary words would not

make previously learned arbitrary words more confusable. Both experiments and computational modelling approaches are well suited to exploring this possibility.

6.4 Overall Conclusions

In this dissertation I have attempted to address the possibility that non-arbitrary associations at the level of the lexicon might be important for language learning.

The pressure for human languages to be both learnable and expressible has been raised with respect to the organization of language at all levels, from phonology (ref) to morphosyntax (Kirby et al., 2015), other than at the level of the lexicon.

However, just as perceptual and cognitive constraints relevant to learning have been proposed to influence the structure of languages generally, I have proposed here that those same constraints exist at the level of individual word-meaning mappings, and thus that they should similarly shape lexical structure.

Although this suggestion is not new, this dissertation has sought to apply a single framework to a wide range of research in psychology and linguistics exploring the task of language learning, and to examine critically how well current theories account for the observed experimental and naturalistic data in the field. The most central contribution of this dissertation to the field generally is the contribution of a parameterization of confusability that might be used to explain differential learnability across a wide range of previous findings. This notion of confusability follows closely from Monaghan et al. (2011) and Gasser (2004) in recognizing that

non-arbitrariness constrains signal space, but goes farther than that by being applicable not only to lexica described as a whole but also to individual learning events. Further, the use of this metric further enhanced my ability to critically explore the possibility that learning non-arbitrary tokens bootstraps the acquisition of later arbitrary tokens.

The effect of the interaction between pressures for languages to be both learnable and expressive is one that has different solutions that are dependent on the size of the signal space and the number and variety of meanings that languages are required to express. Because language learning unfolds over time, with a small initial vocabulary dealing with a rather simple set of words and potentially also a more limited phoneme inventory, the optimal solution for language learning early in development is likely to be different than the optimal solution for language learning later on. The presence of non-arbitrary word-meaning associations, especially in the early-acquired lexicon suggests that languages have been shaped to be learned optimally over the course of development. By taking advantage of benefits for non-arbitrariness when those same non-arbitrary associations do not induce learnability or expressivity penalties, the task of language learning is made easier across the board. The degree to which this is true, and the specific types of non-arbitrary associations leveraged for this purpose, will naturally vary between languages, but future research considering this possibility broadly and linking it to an overarching

theory should help illuminate questions about human cognition and the evolution of language.

References

- Abelin, A. (1999). *Studies in Sound Symbolism*. University Dissertation: Goteborg, Sweden.
- Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign System Studies*, 38, 294-348.
- Akita, K. (2011). Toward a phonosemantic definition of iconic words. In *Semblance and Signification* (Michelucci, P. et., eds), pp. 3-18, John Benjamins
- Alpher, B. (1994). Yir-Yoront ideophones. In *Sound symbolism* (eds L Hinton, J Nichols, JJ Ohala), pp. 161–177. Cambridge, UK: Cambridge University Press.
- Asano, M., Imai, M., Kita, S., Katajo, K., Okada, H., & Thierry, G. (2015) Sound symbolism scaffolds language development in preverbal infants. *Cortex*, 63, 196–205
- Asher, J.E., Lamb, J.A., Brocklebank, D., Cazier, J.B., Maestrini, E., Addis, L., Sen, M., BaronCohen, S., & Monaco, A.P. (2009). A whole-genome scan and fine-mapping linkage study of auditory-visual synesthesia reveals evidence of linkage to chromosomes 2q24, 5q33, 6p12, and 12p12. *The American Journal of Human Genetics*, 84(2), 1–7.

- Assaneo, M.F., Nichols, J.I., & Trevisan, M.A. (2011) The anatomy of onomatopoeia. *PLoS ONE*, 6.
- Aveyard, M. E. (2012). Some consonants sound curvy: effects of sound symbolism on object recognition. *Memory and Cognition*, 40, 83-92.
- Bailey, T.M. & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or Lexical Neighbourhoods? *Journal of Memory and Language*, 44, 568-591.
- Bailey, T.M. & Hahn, U. (2006). Phoneme similarity and confusability. *Journal of Memory and Language*, 52, 347-370.
- Bankieris, K., and Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition*, 136, 186–195.
- Baldwin, D. 1993 Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, 29(5), 832–843.
- Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445-459.
- Bates, D.M., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version.
- Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, 80(2), 290-311.

- Birkbeck, N. (2008). Random shape generator. Retrieved from
<http://www.neilbirkbeck.com/jshape/single.html>.
- Blasi, D.E., Christiansen, M.H., Wichmann, S., Hammarstrom, H. & Stadler, P.F.
 (2014). Sound symbolism and the origins of language. In Cartmill et al. (Eds)
The Evolution of Language: Proceedings of the 10th international conference. Pp.
 391. Singapore: World Scientific
- Blust, R. 1988. *Austronesian root theory: an essay on the limits of morphology*.
 Philadelphia: J. Benjamins.
- Bolinger, D. 1980. *Language; The Loaded Weapon: The Use and Abuse of Language
 Today*. New York: Longman.
- Bremner, A.J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K.J., & Spence, C.
 (2013). “Bouba” and “Kiki” in Namibia? A remote culture makes similar
 shape-sound matches, but different shape-taste matches to Westerners.
Cognition, 126, 165-172.
- Chater, N., & Vitanyi, P. (2003). Simplicity: A unifying principle in cognitive
 science. *Trends in Cognitive Science*, 7, 19–22
- Childs G.T. 1994 African ideophones. In *Sound symbolism* (eds L Hinton, J Nichols,
 JJ Ohala), pp. 178–206. Cambridge, UK: Cambridge University Press.

- Cuskley, C. (2013). *Shared cross-modal associations and the emergence of the lexicon*.
University of Edinburgh Doctoral Thesis. Edinburgh, UK.
- Cuskley, C. and Kirby, S. (2013). Synaesthesia, cross-modality, and language evolution. In: Simner, J. and Hubbard, E.M. (Eds). *Oxford Handbook of Synaesthesia*, pp 869-907.
- Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *British Journal of Psychology*, 52, 259-268.
- Dell, G.S. & Reich, P.A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behaviour*, 20, 611-629.
- De Saussure, F. (1983). *Course in General Linguistics*, Eds. Charles Bally and Albert Sechehaye. Translated by Roy Harris. La Salle, Illinois: Open Court.
- Diffloth, G. (1994) i:big, a:small. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 107-114). Cambridge, MA: Cambridge University Press.
- Dingemanse, M. (2011). Ideophones and the aesthetics of everyday language in a West-African society. *Senses and Society*, 6(1), 77-85.
- Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, 6, 654-672

- Dingemanse, M., Blasi, D.E., Lupyán, G., Christiansen, M.H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19.
- Dingemanse, M. Verhoeft, T., & Roberts, S.G. (2014). The role of iconicity in the cultural evolution of communicative skills. In B. De Boer, & T. Verhoeft (Eds.), *Proceedings of Evolang X: Workshop on Signals, Speech, and Signs*, 11-15.
- Dryer, M.S. & Haspelmath, M. (eds.) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info>, Accessed on 2015-12-17.)
- Dunn, L., & Dunn, L. (1997). *Peabody picture vocabulary test III*. Circle Pines, MN: American Guidance Service.
- Epps, P. 2005. *A grammar of Hup*. PhD dissertation, University of Virginia.
- Farmer, T. A., Christiansen, M. H., and Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, 103, 12203-08

- Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development, 64*, 637-56.
- Fitneva, S., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language, 36*, 967-997.
- Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representatoin: Where might graphical symbol systems come from? *Cognitive Science, 31(6)*, 961-987.
- Gasser, M. (2004). The origins of arbitrariness of language. Proceedings of the 26th Annual Conference of the Cognitive Science Society, pp. 434-439. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gauthier, I., & Tarr, M. (1997). Becoming a 'Greeble' expert: Exploring mechanisms for face recognition. *Vision Research, 37(12)*, 1673-1682.
- Gebels, G. (1969). An investigation of phonetic symbolism in different cultures. *Journal of Verbal Learning and Verbal Behavior, 8*, 310-12
- Grieser, D.L., & Kuhl, P.K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Development Psychology, 24*, 14-20.

- Haiman, J. (1980). The iconicity of grammar: Isomorphism and motivation. *Language*, 56(3), 515–540.
- Haiman, J. (1985) Natural Syntax. Iconicity and Erosion. *Cambridge Studies in Linguistics*, 44. Cambridge: Cambridge University Press.
- Hamano, S. (1998). *The Sound-Symbolic System of Japanese*. Cambridge University Press.
- Harm, M.W., & Seidenberg, M.S. (1999). Reading acquisition, phonology, and dyslexia: Insights from a connectionist model. *Psychological Review*, 106, 491-528.
- Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, 325-336.
- Hinton, L., Nichols, J., & Ohala, J. (1994). *Sound symbolism*. Cambridge, MA: Cambridge University Press.
- Hockett, C.F. (1960). The origin of speech. *Scientific American*, 203, 88-96.
- Huang, Y.H. (1969). Universal magnitude symbolism. *Journal of Verbal Learning and Verbal Behavior*, 8, 155-156.
- Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109, 54-65.

- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B*, 369.
- Johnson, R.C. (1967). Magnitude symbolism of English words. *Journal of Verbal Learning and Behavior*, 6, 508-511.
- Kanera, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How sound symbolism is processed in the brain: A study on Japanese mimetic words. *PLoS ONE*, 9.
- Kantarzis, K., Kita, S., & Imai, M. (2011). Japanese sound symbolism facilitates word learning in English speaking children. *Cognitive Science*, 35, 575-86.
- Kelly, M. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, 99, 349-364.
- Kim K.-O. (1977). Sound symbolism in Korean. *Journal of Linguistics*, 13, 67-75.
- 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 Science*, 105, 10681-10686.
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87-102.

- Klamer, M. (1999). Austronesian expressives and the Lexicon. *Toronto Working Papers in Linguistics*, 16(2). 201–19.
- Köhler, W. (1929). *Gestalt Psychology*. New York, NY: Liveright.
- Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, 114, 19-28.
- Kruspe, N. 2004. *A grammar of Semelai*. Cambridge: Cambridge University Press.
- Landau, B., Smith, L.B., & Jones, S.S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3, 299-321.
- Lindauer, M. (1990). The effects of the physiognomic stimuli taketa and maluma on the meanings of neutral stimuli. *Bulletin of the Psychonomic Society*, 28, 151-154.
- Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioural, developmental, and neuroimaging research into sound symbolism. *Frontiers in Psychology*, 6, 1264.
- Luce, P.A. (1986). *Neighborhoods of Words in the Mental Lexicon*. PhD Dissertation, Dept of Psychology, Indiana University, Bloomington, Indiana.
- Lupyan, G., & Casasanto, D. (2014). Meaningless words promote meaningful categorisation. *Language and Cognition*, 7, 167-193.

- Lupyan, G., Rakison, D.H., & McClelland, J.L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, 18, 1077-1083.
- Maddieson, I. (2013). Consonant Inventories. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/1>, Accessed on 2015-12-17.)
- Maddieson, I. (2013)b. Vowel Quality Inventories. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/2>, Accessed on 2015-12-17.
- Malmberg, B. (1964). Couches primitives de structure phonologique. *Phonetica*, 1, 221-227.
- Maurer, D., Pathman, T., & Mondloch, C.J. (2006). The shape of Boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, 9, 316-322.

- Maurer, D., & Mondloch, C. (2004) Neonatal synesthesia: A re-evaluation. In L. Robertson & N. Sagiv (Eds.), *Synesthesia: Perspectives from cognitive neuroscience* (pp. 193-213). Oxford: Oxford University Press.
- McCune, K. (1983). *The Internal Structure of Indonesian Roots*, Ph.D. dissertation, University of Michigan.
- Miller, G.A. & Nicely, P.E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338-252.
- Monaghan, P., Shillcock, R.C., Christiansen, M.H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B*, 369.
- Monaghan, P., Christiansen, M.H., & Fitneva, S.A. (2011). The Arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, 140, 325-47.
- Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1152-64.
- Newman, S.S. (1933). Further experiments in phonetic symbolism. *American Journal of Psychology*, 45, 53-75.
- Newman, P. (1968). Ideophones from a syntactic point of view. *Journal of West African Languages*, 5, 107-17.

- Newman, P. (2001). Are ideophones really as weird and extra-systematic as linguists make them out to be? In F.K. Erhard-Voeltz, & Christa Kilian-Hatz (Eds.), *Ideophones (pp. 251-8)*. Amsterdam: John Benjamins.
- Newmeyer, F.J. (1993). Iconicity and generative grammar. *Language*, 68, 756-796.
- Nielsen, A.K.S. (2011). *Sound symbolism and the Bouba-Kiki effect: Uniting function and mechanism in the search for language universals*. Master's Thesis: University of Lethbridge, Lethbridge, Alberta.
- Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, 65, 115-124.
- Nielsen, A. & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition*, 4, 115-125
- Nielsen, A., & Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology*, 67, 153-163.
- Nowak, M.A. & Krakauer, D.C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences, USA*, 96, 8028-33.

- Nuckolls, J.B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, 18, 225-252.
- Nygaard, L.C., Cook, A.E., & Namy, L.L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, 112, 181-186.
- Ogura, T. (2006). How the use of ‘non-adult words’ varies as a function of context and children’s linguistic development. In *Studies in language science (5): papers from the fifth annual conference of the Japanese society for language science*, Kobe, Japan, 5–6 July 2003 (eds M NakayamaM Minami, H Morikawa, K Nakamura, H Sirai), pp. 103–120. Tokyo, Japan: Kurosio Publishers.
- Owren, M.J., & Rendall, D. (2001). Sound on the rebound: Returning form and function to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology*, 10(2), 58-71.
- Ozturk, O., Krehm, Madelaine, & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month olds. *Journal of Experimental Child Psychology*, 114, 173-186.
- Pena, M., Mehler, J., & Nespors, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, 22, 1419-21.

- Perlman, M., Dale, D., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, 2.
- Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, 1, 1-15.
- Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369.
- Perry, L.K., Perlman, M., & Lupyan, G. (2015). Iconicity in English vocabulary and its relation to toddlers' word learning. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (Maglio, P. et al., eds).
- Pexman, P.M. & Sidhu, D. M. (2014). Beyond the Bouba/Kiki Effect: The Bob/Kirk Effect. Talk presented at the annual meeting of the Canadian Society for Brain, Behaviour and Cognitive Sciences, Toronto, ON.
- R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramachandran, V.S., & Hubbard, E.M. (2001). Synaesthesia- A window into perception, thought and language. *Journal of Consciousness Studies*, 8, 3-34.

- Rattermann, M.J. & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal mapping task. *Cognitive Development*, 13, 453-478.
- Reilly, J., Westbury, C., Kean, J., & Peele, J. E. (2012). Arbitrary symbolism in natural language revisited: When forms carry meaning. *PLoS ONE*, 7(8), 1-15.
- Revill, K.P., Namy, L.L., DeFife, L.C., & Nygaard, L.C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain and Language*, 128, 18-24.
- Rhodes, R. (1981). On the Semantics of Ojibwa Verbs of Breaking, in W. Cowan, (Ed.) *Papers of the Twelfth Algonquian Conference*, Ottawa: Carleton University Press
- Rouw, R., & Scholte, H. S. (2007). Increased structural connectivity in grapheme–color synesthesia. *Nature Neuroscience*, 10, 792–797.
- Saji, N., & Imai, M. (2013). Goishutoku ni okeru ruizousei no kouka no kentou [the role of iconicity in lexical development]. In *Onomatope kenkyu no shatei–chikadzuku oto to imi (sound symbolism and mimetics)* (eds K Shinohara, R Uno), pp. 151–166. Tokyo, Japan: Hituji Syobo.

- Samuelson, L.K., & Smith, L.B. (2005). They call it like they see it: spontaneous naming and attention to shape. *Developmental Science*, 8, 182-198.
- Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12, 225-239.
- Schultze-Berndt, E. (2001). Ideophone-like characteristics of uninflected predicates in Jaminjung (Australia). In *Ideophones* (eds FKE Voeltz, C Kilian-Hatz).
- Shattuck-Hufnagel, S. & Klatt, D. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, 18, 41-55.
- Silvey, C., Kirby, S., & Smith, K. (2015). Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive Science*, 39, 212-26.
- Simner, J. (2006). Beyond perception: Synaesthesia as a psycholinguistic phenomenon. *Trends in Cognitive Science*, 11, 23-29.
- Simner, J. (2012). Defining synaesthesia. *British Journal of Psychology*, 103, 1-15.
- Tamariz, M., & Kirby, S. (2015) Culture: Copying, compression, and conventionality. *Cognitive Science*, 39, 171-183.
- Taub, S.F. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge: Cambridge University Press.

- Thompson, R.L., Vinson, D.P., Woll, B., & Vigliocco, G. (2012). The road to language learning is iconic: Evidence from british sign language. *Psychological Science*, 23, 1443-48.
- pp. 355–374. Amsterdam, The Netherlands: John Benjamins.
- Thompson, B., Silvey, C., Kirby, S., & Smith, K. (2014). The effect of communication on category structure. In E. A. Cartmill, S. Roberts, H. Lyn, & H. Cornish (Eds.), *The Evolution of Language: Proceedings of the 10th International Conference (EVOlang10)*. Singapore: World Scientific.
- Vinson, D.P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The british signal language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavioural Research Methods*, 40, 1079-1087.
- Watson, R.L. (2001). A comparison of some Southeast Asian ideophones with some African ideophones. *Ideophones*, ed. by. F. K. Erhard Voeltz and Christa Kilian-Hatz, 385–405. Amsterdam: John Benjamins.
- Waxman, S. R., and Hall, D. G. (1993). The development of a linkage between count nouns and object categories: Evidence from 15- to 21-month-old infants. *Child Development*, 64, 1224-1241
- Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-month-old infants. *Cognitive Psychology*, 29, 257-302.

- Wichmann, S., Holman, E.W., & Brown, C.H. (2010). Sound symbolism in basic vocabulary. *Entropy*, 12, 844-858.
- Wicklegren, W.A. (1965). Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America*, 38, 583-588.
- Wicklegren, W.A. (1966). Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America*, 39, 388-398.
- Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, 7, 415-449.
- Yardy, B.J. (2010). *Sound symbolism, sonority, and swearing: An affect induction perspective*. Master's Thesis: University of Lethbridge, Lethbridge, Alberta.
- Yoshida, H. (2012). A cross-linguistic study of sound symbolism in children's verb learning. *Journal of Cognitive Development*, 13, 232-65.

Appendix A- Figures

Animal stimuli from Experiments 1-4



Vehicle Stimuli from Experiments 1-4

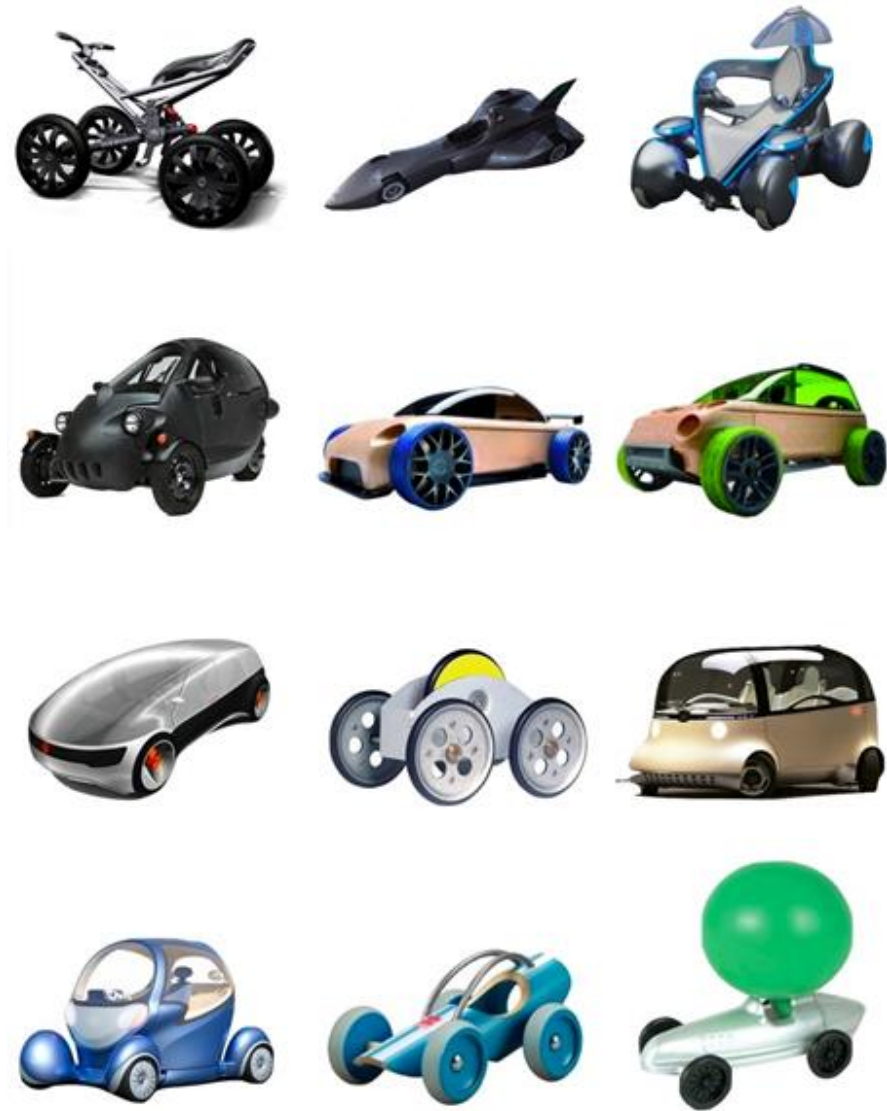
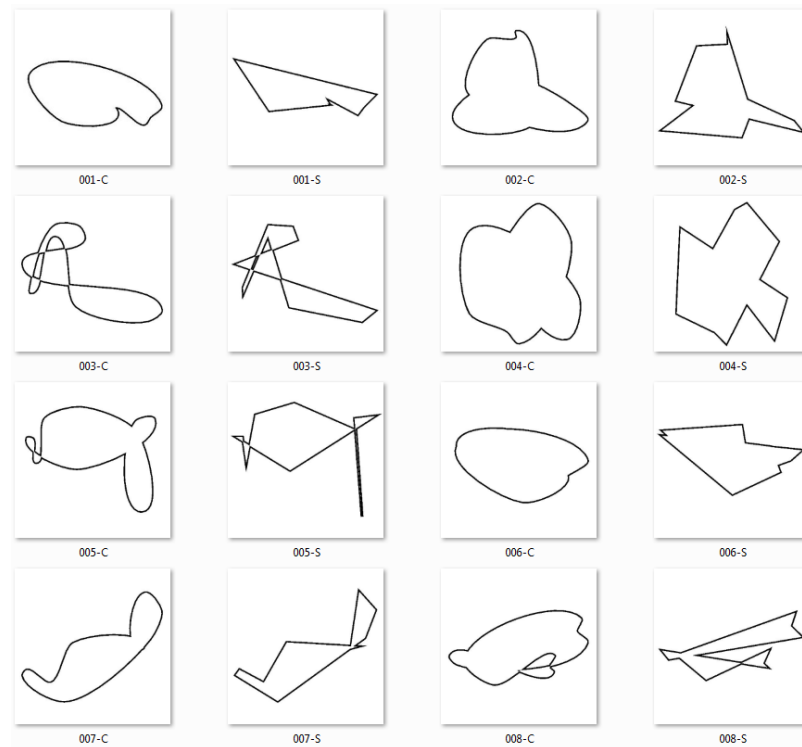
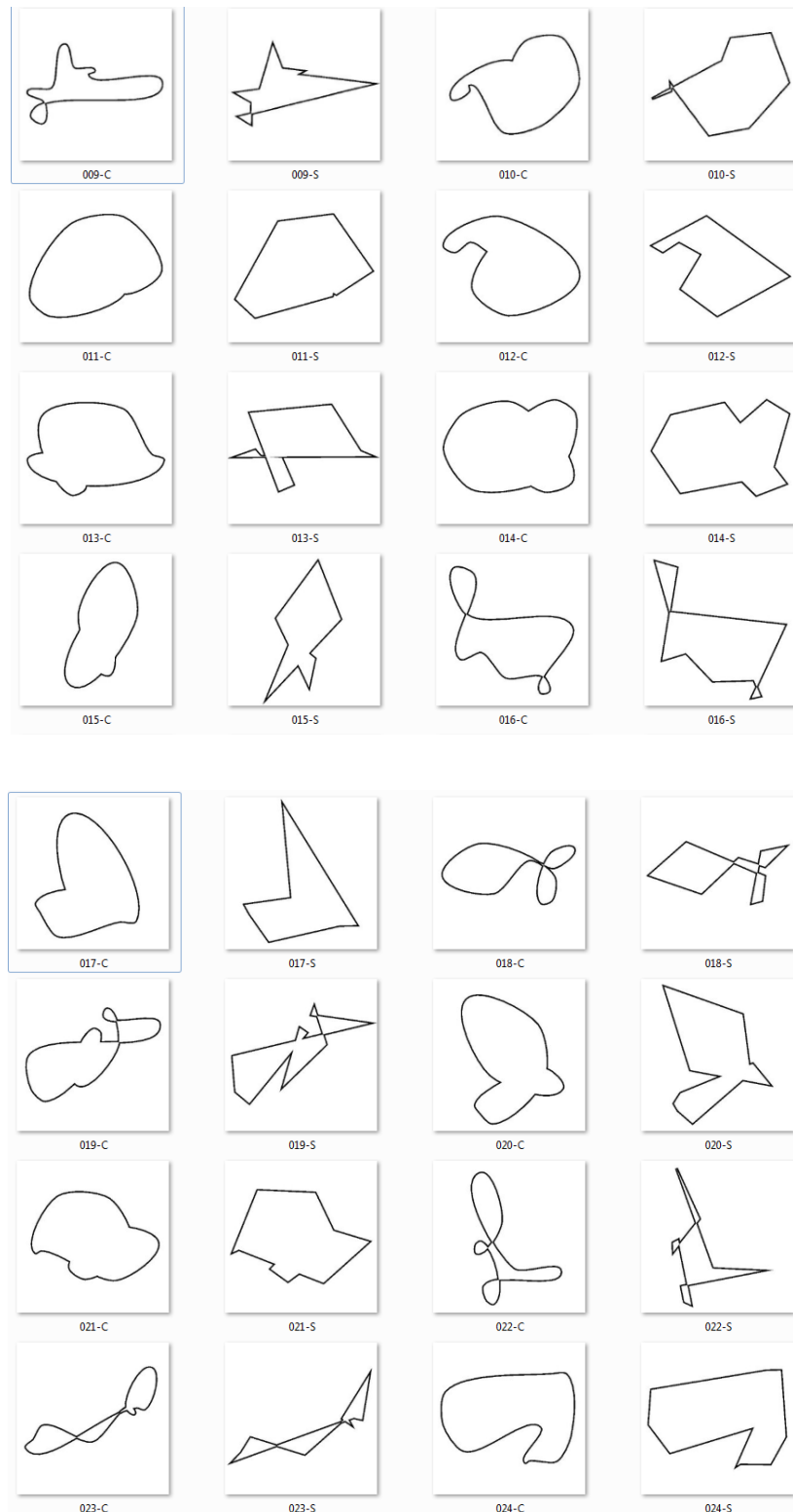
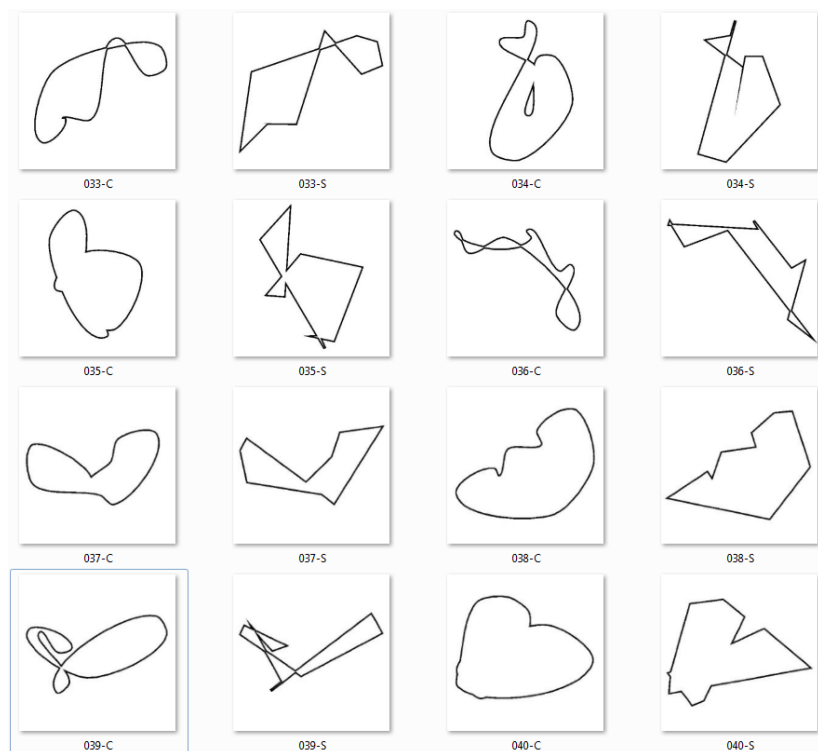
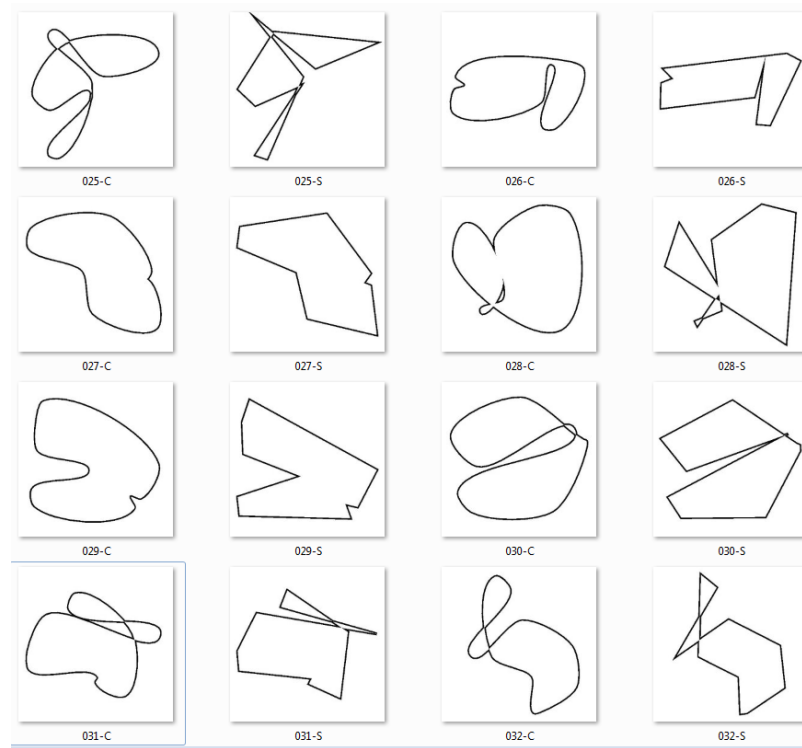
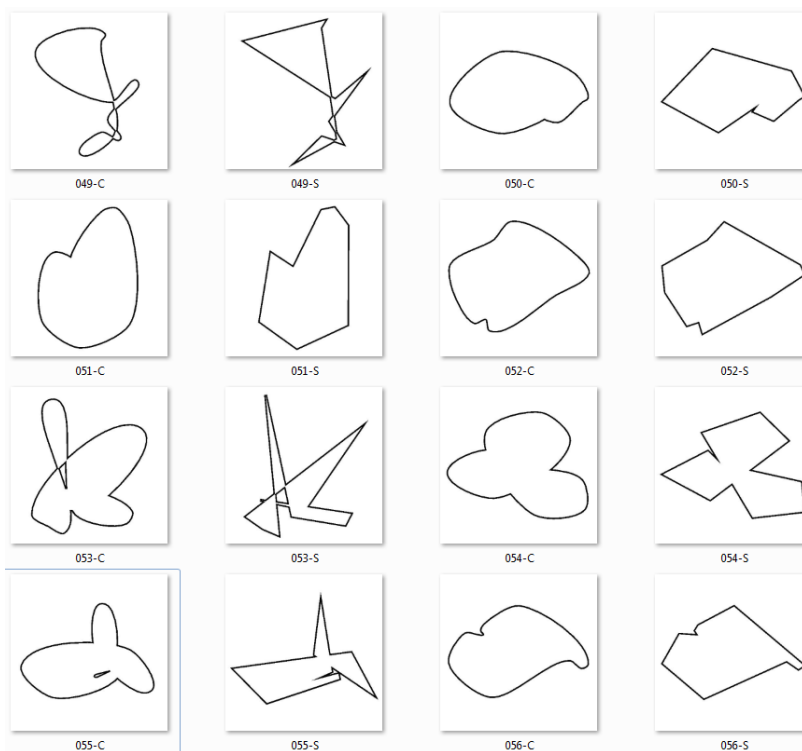
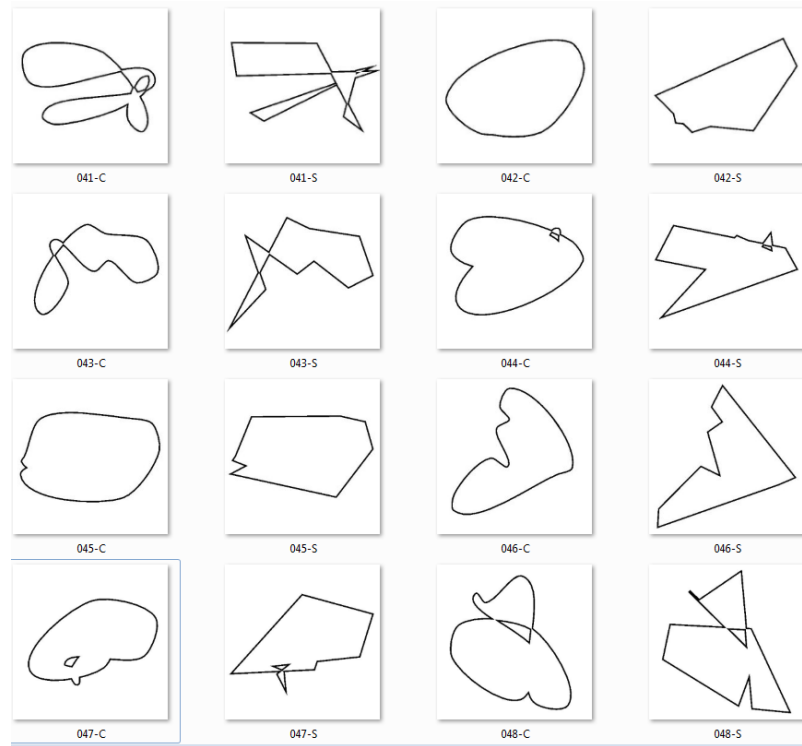


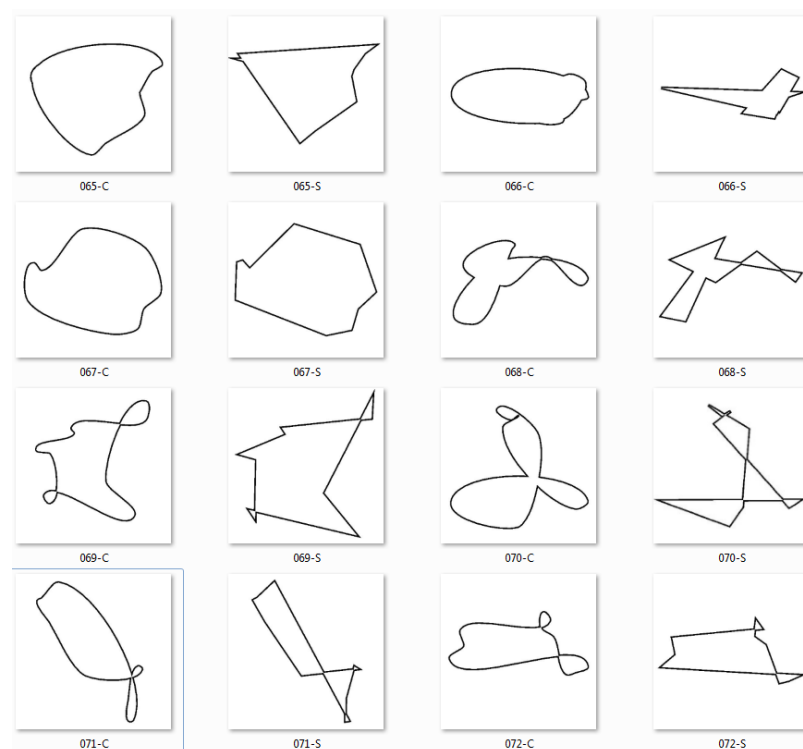
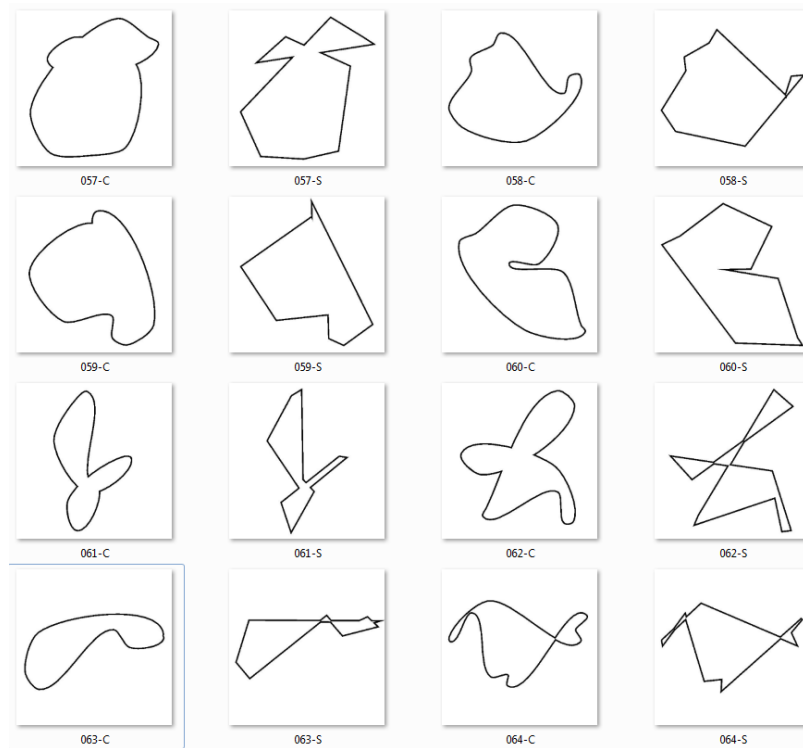
Image Stimuli from Experiment 5

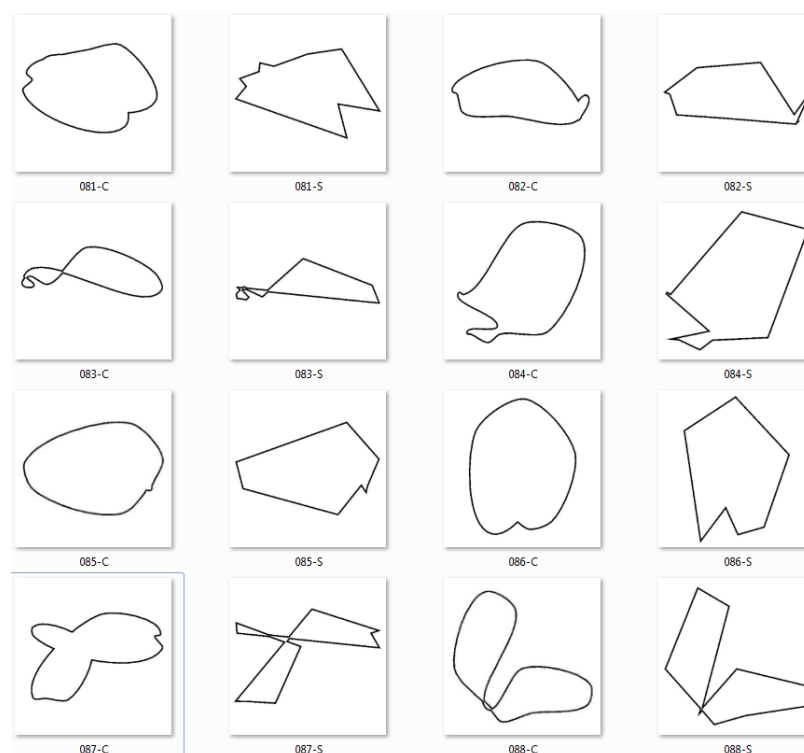
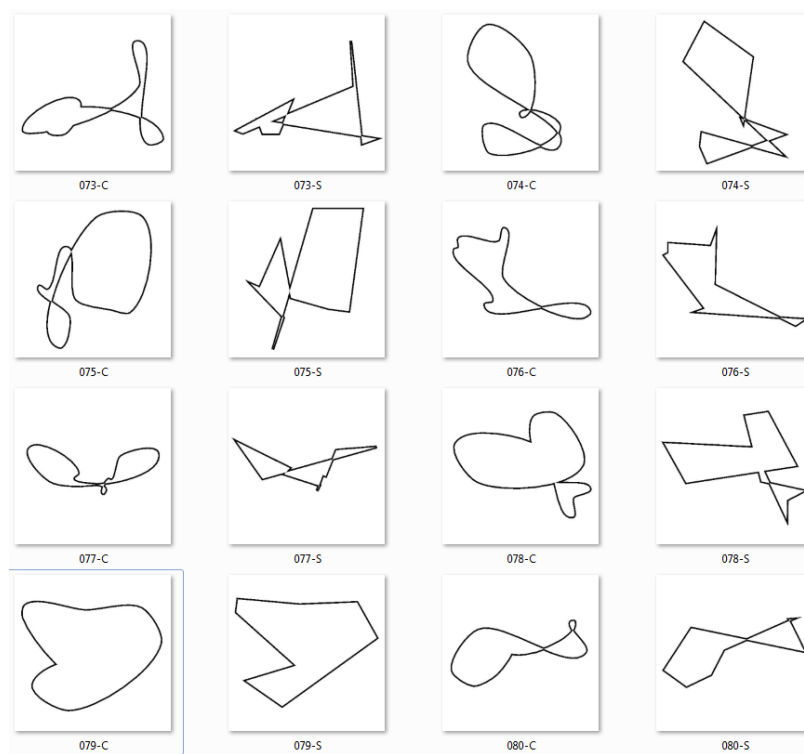












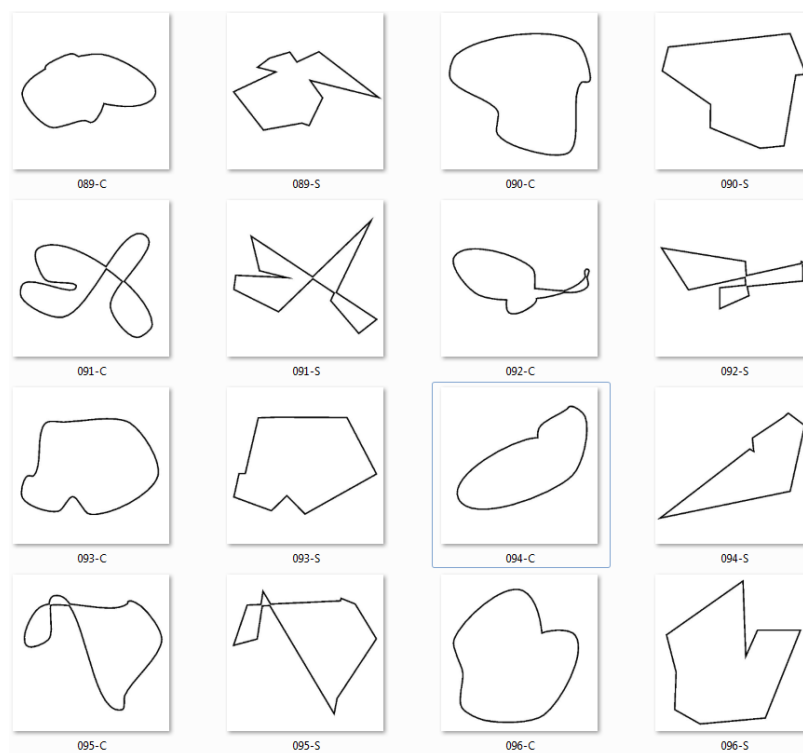
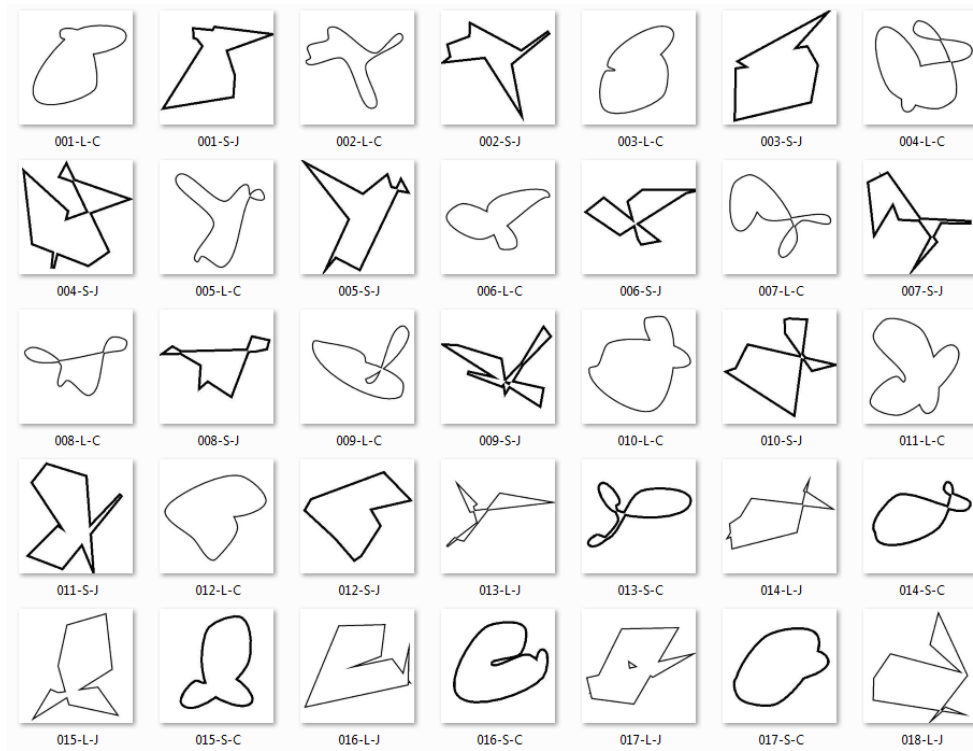
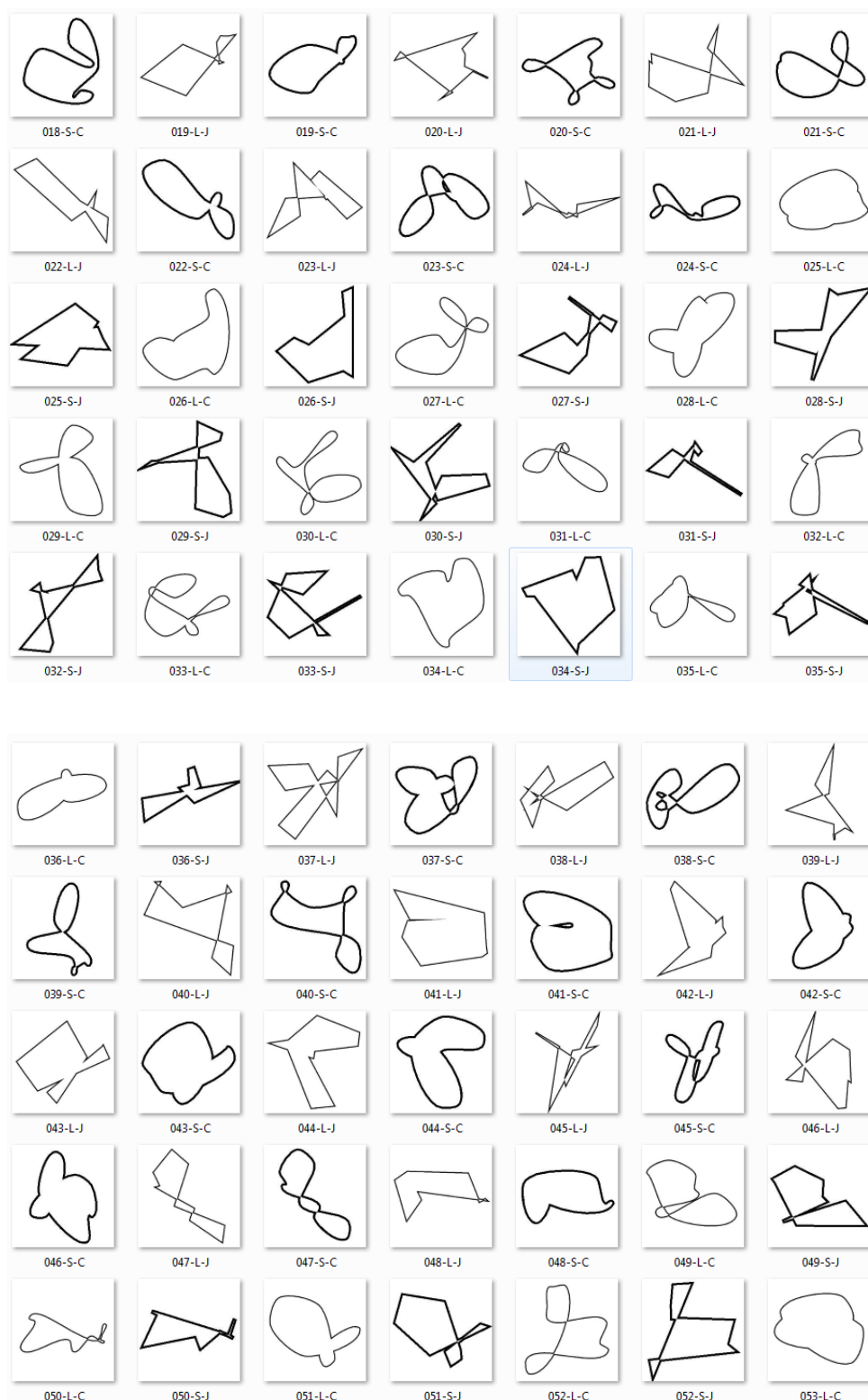
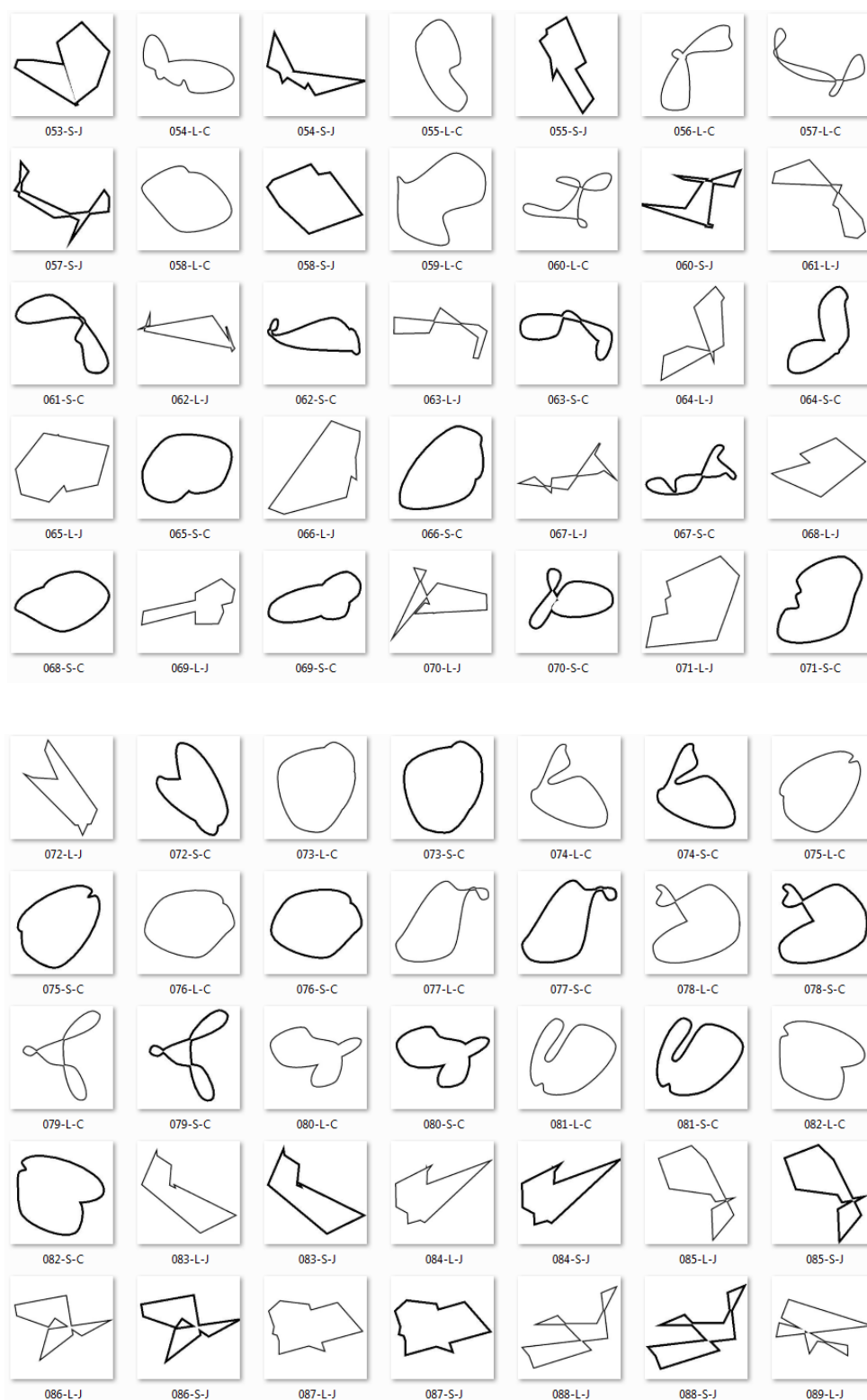
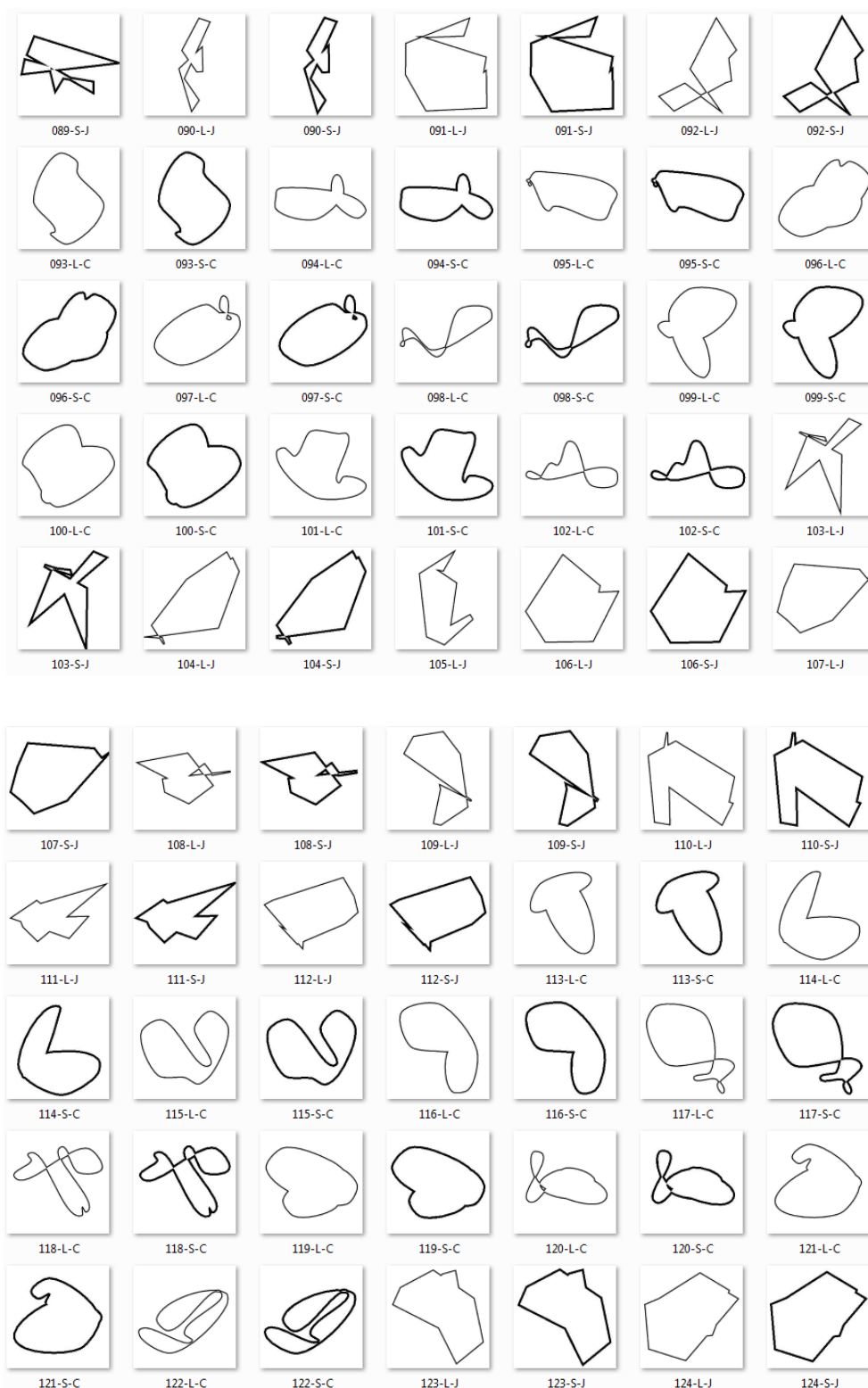


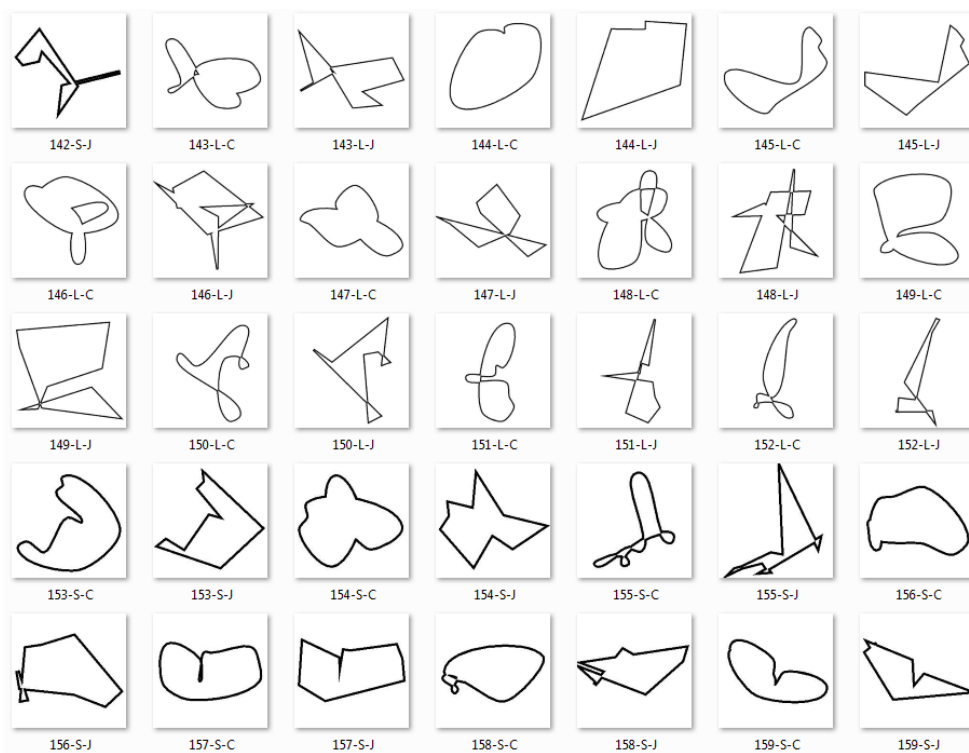
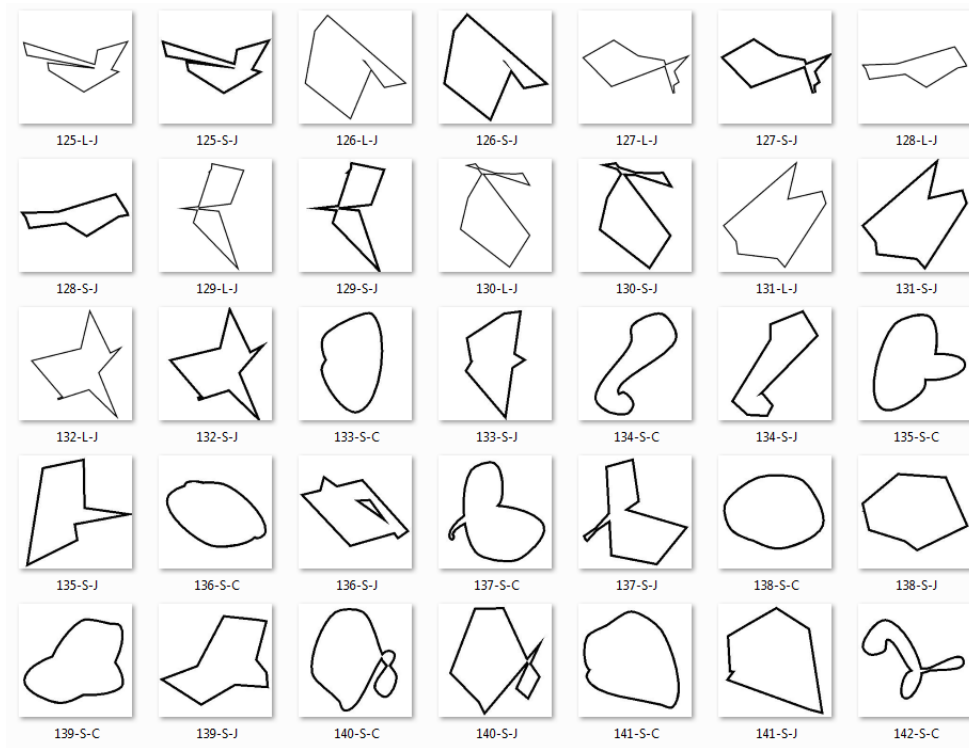
Image Stimuli from Experiment 6











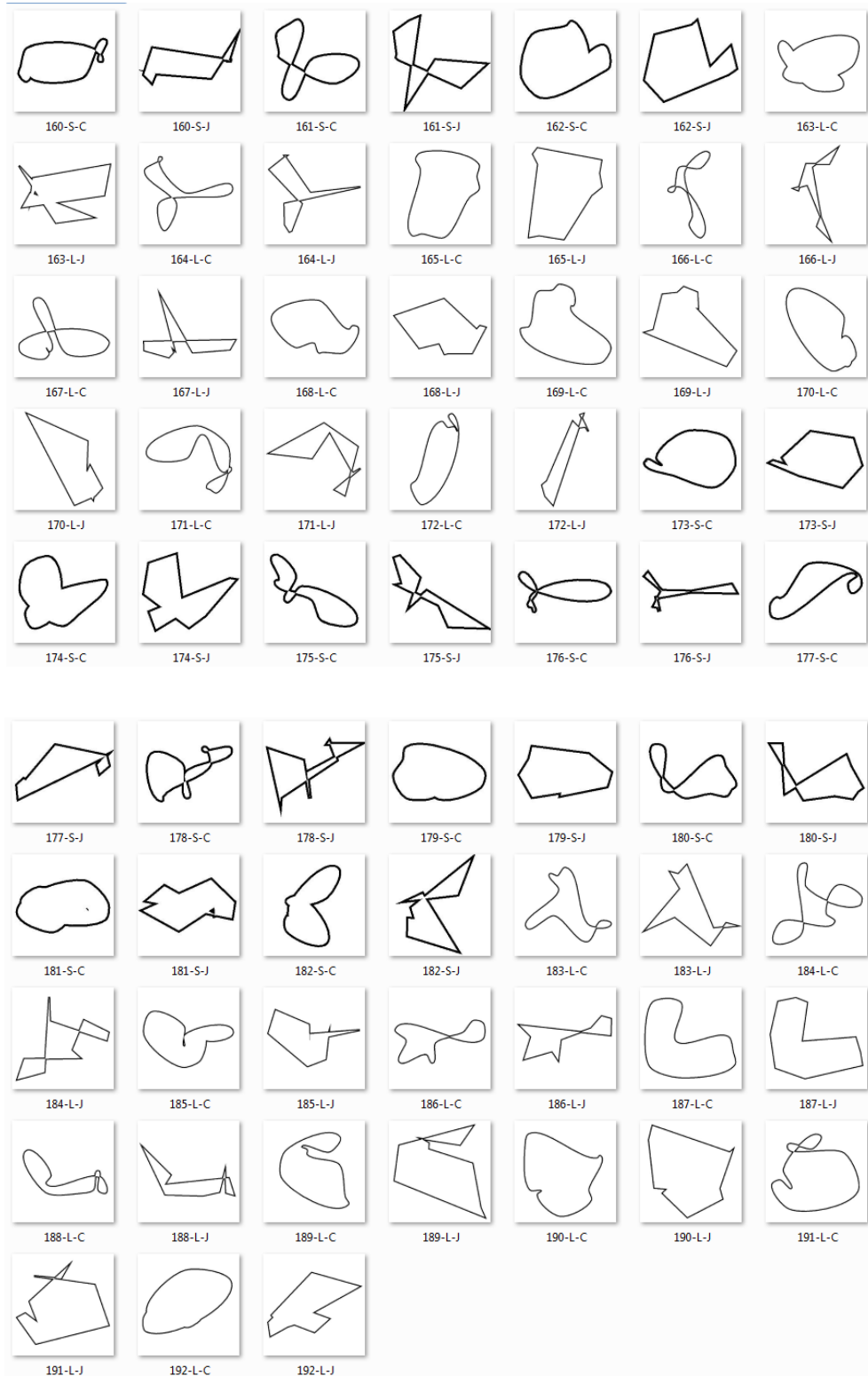


Image Stimuli from Experiment 7

