

8-24-2017

# Conventionalization and Reduction in Natural Language Emergence: An Experimental and Computational Model Investigation

Russell Richie

*University of Connecticut - Storrs*, [russell.richie@uconn.edu](mailto:russell.richie@uconn.edu)

Follow this and additional works at: <https://opencommons.uconn.edu/dissertations>

---

## Recommended Citation

Richie, Russell, "Conventionalization and Reduction in Natural Language Emergence: An Experimental and Computational Model Investigation" (2017). *Doctoral Dissertations*. 1541.  
<https://opencommons.uconn.edu/dissertations/1541>

Russell Richie, PhD – University of Connecticut, 2017

Conventionalization and Reduction in Natural Language Emergence: An Experimental and  
Computational Model Investigation

Russell Richie, Ph D.

University of Connecticut, 2017

In the emergence of natural languages, two processes typically co-occur: new form-meaning mappings are being conventionalized, and these forms reduce, i.e., shorten, in various ways. Despite this oft-noted co-occurrence, and the existence of theories of conventionalization and reduction *separately*, there are no explicit, mechanistic, and empirically-motivated theories of the connection between the two processes. To fill this gap, the present dissertation aims to better understand this relationship both empirically and theoretically, with experimental semiotics and agent-based modeling.

In the experiment, eight groups of four hearing, non-signing English-speaking participants were brought into the lab. Quads split into rotating dyads which took turns using gestures – usually highly iconic – to communicate (images of) basic objects (e.g., cow, orange, boy) to each other. As in naturalistic language emergence, quads conventionalized, reduced, and improved success at the task in a self-organizing fashion. Across objects, slope of conventionalization negatively correlated with slope of reduction, suggesting that those objects that conventionalize more also reduced more. Critically, we found that participants reduced *more* after communication failure, inconsistent with a listener-oriented/rational-agent model of the link between conventionalization and reduction, whereby conventionalization causes communication success, which causes agents to try reducing their utterances (i.e., to try to save effort when they think communication is likely to succeed).

To develop a new theory of the link between conventionalization and reduction, we implemented an agent-based model of gesture production, comprehension, and learning. Three key properties subserve the model: (1) probabilistic gesture-object mappings initially set to model iconicity (which participants in the experiment spontaneously use), (2) language production dependent on a notion of informativeness (understanding one's own utterance), and (3) a single learning mechanism in listeners, whereby listeners align their probabilistic associations with speaker's utterances. This model simultaneously conventionalized and reduced, and captured other aspects of our experimental setting.

Consistent with previous naturalistic and experimental work, our experiment and model each show parallel conventionalization and reduction. To our knowledge, our model is the first to simultaneously capture these two phenomena, and shows that conventionalization and reduction can simultaneously emerge from independently necessary mechanisms of language production, comprehension, and learning, rather than from agents rationally selecting optimal lexicons, or optimally responding to their interlocutors' state of understanding. With our work as a starting point, future work could further investigate how conventionalization and reduction relate to the emergence of phonology and grammaticalization.

Conventionalization and Reduction in Natural Language Emergence: An Experimental and  
Computational Model Investigation

David Russell Mahood Richie

B.A., Bates College, 2009

M.A., University of Connecticut, 2013

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2017

Copyright by  
David Russell Mahood Richie

APPROVAL PAGE

Doctor of Philosophy Dissertation

Conventionalization and Reduction in Natural Language Emergence: An Experimental and  
Computational Model Investigation

Presented by

David Russell Mahood Richie, B.A., M.A.

Major Advisor

---

Marie E. Coppola

Co-Major Advisor

---

Whitney Tabor

Associate Advisor

---

Adam Sheya

University of Connecticut

2017

## Acknowledgments

I have many people to thank, and much to be thankful for. Because I am only here today, in science and in life, because of the gifts from those around me.

I owe the most to my loving, devoted family. My parents **David and Dail Richie** were the first to nurture my curiosity, reading to me many nights as a child, and stimulating my interest in the arts (which is a short step from the sciences!). My sisters **Julia Richie Sammin** and **Carolyn Richie Tavares** constantly inspire me while making me laugh. And they probably provided rich linguistic input to their little brother!

My academic journey has owed much to brilliant and kind mentors and colleagues as well. At Bates College I was fortunate enough to be inspired into doing psychological research by **Drs. Kathy Mathis, Todd Kahan, Kathy Low, Helen Boucher**, and many others. Kathy Mathis and Todd Kahan patiently supervised my senior thesis on verb learning. I thus largely have them to thank for my enduring interest in language. That same thesis landed me a post-bac research assistant job with **Dr. Kathy Hirsh-Pasek** at Temple University. Kathy, as well as **Drs. Nora Newcombe** (also of Temple) and **Roberta Golinkoff** (of University of Delaware), were instrumental in introducing me to the wider world of cognitive science, language acquisition, and psycholinguistics. It was in that world – particularly the Boston University Conference on Language Development in 2010 – that I was able to meet **Dr. Marie Coppola**. I had read of Marie and her colleague's work on Nicaraguan Sign Language and homesign back at Bates, and had been immediately fascinated with it. That fascination, which continues to this day, ultimately led me to University of Connecticut to study language emergence with Marie.

Thanking Marie really deserves much more than a paragraph and a few words. To say the least, she is brilliant, kind, hard-working, fierce, funny. She helped keep me going at my lowest

point. Without her, graduate school would have been a much, much rougher ride, and I may not have finished at all.

As if having Marie as an advisor wasn't enough fortune for one person, I've been lucky enough to find a second amazing mentor at UConn. I quickly discovered during my first year the intellectual and personal common ground I shared with **Dr. Whitney Tabor**. Patient and generous with his time, after a couple years Whit agreed to co-advise me with Marie. Having both Whit and Marie as advisors has truly been a blessing: they've made me a better scientist, thinker, and person.

Many other faculty at UCONN were instrumental to my development. **Dr. Adam Sheya**, my committee member and seminar professor, has always given me time to talk about stats, career, and life. **Dr. Letitia Naigles** has commented on countless of my papers, presentations, and posters over the years. **Dr. James Magnuson** invested in me financially, through the **UConn IGERT/ Neurobiology of Language program**, and with his comments on my work. Thanks to **Dr. Paul Alloppenna** for shooting the breeze about science, politics, and everything in between. **Dr. Stefan Kaufmann** began my education and career in computational linguistics, which I hope (and plan!) to continue after University of Connecticut. Thank you to **Drs. Diane Lillo-Martin, William Snyder, Rachel Theodore, Austen Clark, Carl Coelho** and **Thomas Bontly** for allowing me to be a part of teaching cognitive science.

UConn also introduced me to many other amazing colleagues: my collaborators **Dr. Charles Yang, Dr. Matt Hall, Dr. Kaja Jasinska, Dr. Sandra Wood, Dr. Diane Lillo-Martin**; my labmates **Dr. Emily Carrigan, Dr. Deanna Gagne, and Jessica Contreras**; my research assistants Jenna Shaw, Shea Marsella, Sarah Lodge, and Megan Brown; my friends **Katie Shaw, Dr. Amanda Rendall and Peter Perrino, Garrett Smith and Pam Fuhmeister**,

**Josh Green, Parker Tichko, Dr. Kaitlyn Flannery;** and many, many others. You all made the journey fun, interesting, and, well, doable!

One UCONN student colleague deserves special mention: **Dr. Iris Chin.** I have had more fun, growth, and love with you than I thought possible. I can't wait to take the next step in our lives together.

Finally, I want to give thanks to all the people that have participated in my and my colleagues' research over the years. Many of our participants were undergrads at UCONN doing studies for course credit. I thank these students – I hope their experience was educational! However, I want to express my deepest appreciation for the Deaf persons in the US and Nicaragua that I have worked with over the years. Thank you for sharing your language and community with me.

I am incredibly lucky to have all of you in my personal and professional lives. If everyone had people as amazing as you in their lives, the world would be a better place.

This research was supported by NSF IGERT Grant 1144399 to the University of Connecticut and NSF INSPIRE 1246920 to the University of Connecticut.

## Table of Contents

### Table of Contents

<b>Chapter 1. Introduction.....</b>	<b>1</b>
The problem of language emergence.....	1
Psycholinguistic and developmental interest in language emergence .....	1
Narrowing into lexicon emergence .....	2
Phenomenon One: Conventionalization .....	2
Phenomenon Two: Reduction .....	3
Broader implications of conventionalization and reduction .....	6
Theories of conventionalization and reduction.....	9
Theories of the <i>relationship between</i> conventionalization and reduction .....	13
<b>Chapter 2 - Experimental Method .....</b>	<b>15</b>
Participants .....	16
Design & Procedure .....	17
Materials .....	20
Coding.....	20
Measuring string similarity .....	21
<b>Chapter 3 - Experimental Results .....</b>	<b>22</b>
Experimental data have ecological validity .....	22
Different quads conventionalize different systems, with some arbitrariness, in a self-organizing manner.....	24
Conventionalization and reduction are correlated .....	30
Analyzing the relationship between reduction and communication success .....	32
<b>Chapter 4 - Computational Modeling .....</b>	<b>33</b>
Key aspects of the experiment to be modeled .....	34
A previous model of conventionalization of referring expressions in an emerging signed system .....	36
A new model of conventionalization and reduction.....	38
Model behavior.....	44
Conclusions of modeling.....	50
<b>Chapter 5 - Discussion.....</b>	<b>52</b>
Relation to existing theories of language emergence, conventionalization and reduction .....	54
Limitations of the present work .....	60
Future directions.....	64
<b>Conclusion.....</b>	<b>68</b>
<b>References .....</b>	<b>70</b>
<b>Appendix A. Experimental items.....</b>	<b>79</b>

## Chapter 1. Introduction

### **The problem of language emergence**

Most scholars considering how an individual could come to know a language presume that the individual is being exposed to a full, natural language produced by elder individuals -- in other words, the learner is largely recapitulating the language they were exposed to. But where did those elders get their language? To avoid infinite regress, we must imagine that, at some point, a new language was bootstrapped *de novo*, presumably when individuals had need to coordinate with each other. Better understanding how such *de novo* language creation could result from individual events of human coordination is the goal of this dissertation.

### **Psycholinguistic and developmental interest in language emergence**

Before exploring language emergence further, it is worth briefly motivating an interest in the phenomenon from developmental and psycholinguistic perspectives. While language emergence is a problem of co-creating among *groups* of individuals, the nature of the *individual* is still critical. That is, language emergence must rely on individuals representing individual concepts and propositions to be conveyed, and encoding and decoding these into/from words and sentences, all in real time. These processes are, of course, the domain of psycholinguistic theorizing. Further, if such individual communicative/coordinative events are to lead to a long-lasting, shared language system, then individuals can not ‘reinvent the wheel’ with every new coordinative event -- in other words, they must *learn* from each event and alter their future linguistic behavior accordingly. How exactly that happens is naturally of interest to developmentalists [for example, see work by, i.a., Senghas, Kita, and Ozyurek (2004) and Senghas and Coppola (2001) for investigations of how the nature of the learner impacts language emergence].

## Narrowing into lexicon emergence

We now refine our interest within language emergence. There are many aspects of language that must be created, from phonetics/phonology up to syntax (and perhaps beyond to semantics and pragmatics). In this work, we start with (but will occasionally branch out from) a focus on *lexicon*<sup>1</sup> emergence. The lexicon is of particular interest in language emergence, as it is arguably the most fundamental but still meaningful and communicative level of language organization (cf. phonological units which are meaningless; prosodic affective cues which are often considered para-linguistic; or syntactic and semantic phrases which are not fundamental but rather composed). In particular, we will investigate two processes occurring in lexicon emergence:

### Phenomenon One: Conventionalization

First, we examine the emergence of conventional (i.e., shared across a population) form-meaning mappings. We will consider a form-meaning mapping ‘conventional’ if (1) it is shared across a population, and (2) equally good alternatives exist (one can say ‘dog’ or ‘hund’) and the choice among these is in some sense ‘arbitrary’<sup>2</sup>. This is naturally a critical problem in language emergence as knowledge of conventions is usually thought to be how one understands my utterance of ‘dog’ or ‘hund’ to refer to a kind of four-legged, furry animal (Hockett, 1960). How

---

<sup>1</sup> Here we use lexicon to mean an inventory of form-meaning associations, and do not intend to invoke all the complexity that a mature language’s lexicon might have. For example, we do not presume that the form-meaning associations we study are ‘words’, or have syntactic features like grammatical category. It may be best to think of our object of study as merely ‘referring expressions’.

<sup>2</sup> There is much debate about what makes a behavior conventional, e.g., whether conventions require common knowledge (this is a technical sense of ‘common knowledge’ attributed to philosopher David Lewis’s (1969) *Convention*; roughly,  $p$  is common knowledge if and only if everyone knows  $p$ , everyone knows that everyone knows  $p$ , everyone knows that everyone knows that everyone knows that  $p$ , etc.). I will not assume common knowledge, and will, for the most part, be abstracting away from other philosophical debates about conventions (e.g., (1) whether a formerly popular convention that is no longer followed, like sending thank you notes, still constitutes a convention, and (2) whether users must *know* that alternative conventions [like ‘perro’ and ‘hund’] exist). At the same time, I leave open the possibility that this dissertation and later work based on it could relate to issues of philosophical concern. The Stanford Encyclopedia of Philosophy entry on conventions (Rescorla, 2015) is an excellent, accessible introduction to such a philosophical treatment of conventions.

do such conventional lexicons arise, and how do individual events of language use and learning shape lexicon emergence? One can imagine that, in the beginning of language emergence, different people could communicate a single meaning (say, ‘cow’) using different forms. One person could make a gesture iconically showing MILKING and another could make a gesture indicating HORNS. At this point, iconicity, not convention, would be the basis of successful communication. However, this iconicity can bootstrap conventionalization; because interlocutors can understand each other’s references to some degree, they have a way to bootstrap their development of a precise form-meaning map. Thus, when person *A* utters MILKING, person *B* understands *A* means ‘cow’ (when ‘cow’ is plausible and other possible meanings associated with MILKING, like ‘I want milk’, are not)<sup>3</sup>. Further, even if *B* typically uses HORNS, *B* can update their lexical entry for ‘cow’ to MILKING, thus achieving conventionalization between this pair. The present experimental and modeling work in this dissertation will rely on iconicity to bootstrap conventionalization and language emergence in this way.<sup>4</sup>

## **Phenomenon Two: Reduction**

The second process that we investigate is less obvious *a priori*, but likely important nonetheless. This is the process of *reduction*, whereby the forms of linguistic elements – from phrases down to morphemes -- are either eroded, in terms of properties like form duration or reach of intended articulatory targets, or entirely eliminated, leading to fewer discrete elements like phonemes or morphemes in a linguistic unit. This is an often-noticed phenomenon in both

---

<sup>3</sup> We are abstracting away from a very important problem often noted in word learning in ontogeny, but which also plausibly affects language emergence: the gavagai problem (Quine, 1960). That is, an utterance of MILKING could plausibly mean ‘cow’, ‘let’s get some milk’, ‘this is what we do with the animal under discussion’, and infinitely many more meanings.

<sup>4</sup> Of course, it is a debate the extent to which different communication mediums afford iconicity, and thus the extent to which iconicity could bootstrap language emergence and conventionalization in different mediums. The gestural-visual modality seems to afford more iconicity than the vocal-auditory modality. We will return to this issue in the discussion.

naturalistic and experimental investigations of language emergence and change. To give a few examples, (1) for most English words starting with ‘kn’, like ‘knight’, ‘knock’, or ‘knob’, we speakers of Modern English do not pronounce aloud the ‘k’, but speakers of Old English did – the ‘k’ has since been entirely reduced, (2) the contemporary monomorphemic, monosyllabic American Sign Language sign for ‘tomato’ condensed from an historically early compound of the signs for ‘red’ and ‘slice’ (Frishberg, 1975), and (3) present day colloquial ‘gonna’ (e.g., “I’m *gonna* walk the dog”) is derived from ‘going to’ (Perez, 1990). Even in a short communicative setting (e.g., a radio show; Fowler and Housum, 1987), subsequent uses of a form get reduced. For a comprehensive review of reduction, see Jaeger and Buz (2016). This set of examples, then, illustrates that reduction is a robust phenomenon, affecting both morphological and phonological/phonetic units, over short and long timescales, in early and late stages of language evolution, and in multiple language modalities (sign and speech). One further important quality of the phenomenon of reduction is especially worth commenting on: the linguistic variables affected by reduction can be either *continuous* [e.g., the duration of the schwa vowel (the ‘o’) in ‘memory’] or *discrete* (e.g., the number of syllables in ‘memory’, which could be 2 or 3, depending on whether one completely eliminates the ‘o’). Continuity vs. discreteness is an important distinction that we’ll make throughout the present work so that we don’t conflate different phenomena and theories of those phenomena, but it is not going to be a primary focus of this work, and so we will not say too much new about it. We only mention it so the reader has a clear idea of our interest: we will be focusing on cases of morpheme elimination (a case of discrete reduction).

We see similar reduction phenomena even early in language emergence, like those found by Meir et al. (2010), which deserve special attention here. They investigated the emergence of

compounds (in English, this includes things like “dog house”) in the emerging Al-Sayyid Bedouin Sign Language in the Negev Desert of Israel. They found that as compounds became conventionalized in the community, they also became more reduced. In particular, they observed that in less conventionalized compounds in ABSL, signers tended to use multiple signs (on the order of two to five) in sequence to describe an object. For more conventionalized compounds, however, such sequences were reduced to two or three signs. Similarly, older signers often used compounds where younger signers produced a single sign. For example, in one group of signers, the older cohort produced a compound MOVIE^WIDE-OBJECT to describe a ‘TV set’, whereas the younger cohort in that group only signed MOVIE. In another group, the two older signers produced compounds for ‘closet’ (CLOTHES^DOORS) and for ‘dove’ (PECK^WINGS), while their three daughters produced just DOORS and WINGS. Once again, note that this reduction is discrete, concerning *count* (0, 1, 2, 3, etc.) of what are essentially morphemes. Osugi, Supalla, and Webb (1999) likewise found that, for a homesign-using family on Amami Island off the coast of Japan, the conventionalization of a compound for the concept ‘year’ occurred as a four-gesture compound evolved into a two-gesture compound. The same phenomenon arises in Nicaraguan Sign Language (NSL). Multiple Nicaraguan homesigning families, who are thought to approximate the pre-NSL state, use unconventional multimorphemic expressions for objects that the Nicaraguan Sign Language community describes with highly conventionalized, monomorphemic expressions (Richie, Yang, Coppola, 2014; Richie, unpublished data). *It is this co-occurrence between conventionalization and reduction in naturalistic cases of language emergence that is the central phenomenon of interest in this dissertation.*

Co-occurrence between conventionalization and reduction can also be observed in experimental work. In one study in a now classic line of work, Schober and Clark (1989) had

pairs of participants tell each other, in English, how to arrange 12 complex figures (“Tangrams”). They found that, as participants came to agree on referring expressions for individual figures and became more accurate at the task, their referring expressions got shorter (a case of morpheme reduction, rather than phonetic/phonological reduction; notice, too, that this reduction is *discrete* rather than continuous). For example, one tangram was initially described as “like a, a dancer or something really weird. Um, and, has a square head”. At the end of the experiment, the same tangram was described as “the dancer with the big fat leg”, an utterance half as long as the initial. More recently, Caldwell and Smith (2012) asked groups of participants to describe colors to each other purely through drawing, and found that, as success and conventionality in groups improved, the drawings became simpler and more reduced. Similarly, Namboodiripad, Lenzen, Lepic and Verhoef (2016) asked hearing non-signing participants to play silent, gesture ‘charade’-like games, and observed not only conventionalization,<sup>5</sup> but also reduction in the size of the articulatory space and in the distance traveled by the articulators in that space (continuous variables).

### **Broader implications of conventionalization and reduction**

Before exploring how and why conventionalization and reduction happen the way they do (and particularly why they co-occur), we consider additional reasons why conventionalization and reduction are of broader theoretical interest.

Linguists, recognizing Hockett (1960)’s Duality of Patterning, often divide language into two broad domains of structure: the P-side, which concerns phonological, prosodic, and phonetic properties of language, and the S-side, which concerns syntactic and semantic (and usually pragmatic) properties of language. To simplify quite a bit, the P-side concerns the medium of

---

<sup>5</sup> Namboodiripad et al.’s (2016) analyses actually measure communication success, but they informally report conventionalization as correlating with reduction.

language (speech, sign, writing, etc.), and the S-side concerns meaning. Each side has a very elaborate computational nature that is expressively powerful (see Heinz and Idsardi, 2013 for a comparison of the computational complexity of S-side structures and P-side structures in terms of the Chomsky Hierarchy) and also particularly efficient for communication (for example, see Nowak & Krakauer, 1999 for arguments that phonology is an efficient solution to the problem of noisy communication). These are thus clearly important properties of our language system, and their emergence begs for explanation. We will make the case now that conventionalization and reduction may play a role in the development of both S-side complexity and P-side complexity.

First, concerning the P-side, conventionalization and reduction may play a role in the emergence of phonological combinatoriality, the property of language where morphemes are composed of discrete, meaningless elements. For example, Sandler, Aronoff, Meir, and Padden (2011), using data from the emerging Al-Sayyid Bedouin Sign Language (ABSL), make the following argument. They propose that, early in language emergence, iconicity is the basis of communication and thus controls language form. Only when conventionalization takes root and shared convention forms the basis of communication are forms freed from the requirements of iconicity (and attention to the iconic form-meaning mapping is reduced). They are then allowed to reorganize (under pressures for, e.g., formal symmetry and ease of articulation) into a phonological system with a limited inventory of meaningless (here, manual) elements that combine in constrained ways (see Brentari, 2007, 2017 for discussion of sign language phonology). For example, a gesture for ‘tree’ may initially iconically represent a tree, with the arm representing the trunk and the hand the canopy. Over time, users may recognize the gesture as conventionally referring to ‘tree’, completely short-circuiting the role of iconicity, and allowing the form to change in ways that erode its iconicity while also bringing the form into

closer relation with other signs/gestures (perhaps the hand-shape in the tree gesture becomes more aligned with phonetically similar hand-shapes in other signs, leading to a more proper discrete phonemic inventory). Consistent with this, Roberts, Lewandowski, and Galantucci (2015) and Verhoef, Kirby, and de Boer (2015), which both manipulated the availability of iconic form-meaning mappings to participants, provide experimental evidence that iconicity impedes phonological combinatoriality in emerging communication systems. Reduction -- also often due to redundancy of form, as redundant elements can be eliminated without reducing a form's identifiability -- similarly erodes iconicity, as parts of a form that contribute to iconicity are reduced or eliminated entirely, again enabling the transition to phonological systematicity. For example, returning to the 'tree' example from above, reductive forces may eliminate, say, a hand-shaking movement that iconically resembles a tree's leaves blowing in the wind, further freeing the gesture from iconicity and allowing it to undergo the phonologizing changes described above. Thus, an accurate empirical and mechanistic picture of conventionalization and reduction may be necessary to understand how phonological combinatoriality emerges.

Second, conventionalization might also be involved in an important S-side process: grammaticalization, or the process by which lexical elements (like nouns and verbs) become grammatical ones (like tense or conditional markers) over historical time. Grammaticalization is particularly important and of very general interest -- for not only language emergence specialists but also language scholars of all stripes -- as it is an often-invoked explanation for how languages acquire their abstract, flexible expressive power. For example, 'will' can now be used as an auxiliary, grammatical marker of futurity, whereas in Old English the ancestral form 'willan' was a main verb meaning 'to want/wish'. It has been argued that a form's grammaticalization is driven by its frequency (Bybee, 2003; 2006). Thus, once a form is conventional, it will increase

in frequency, and thus grammaticalized, conventionalization may cause grammaticalization and concomitant grammatical sophistication in emerging and changing languages. It may also be that more sophisticated, grammaticalized structures can only emerge after more basic ones have been conventionalized in a population. Such a pattern appears to obtain in Nicaraguan Sign Language. Kocab, Pyers, and Senghas (2015) found that while First- and Second-Cohort NSL users utilize the left-right axis of signing space to refer to objects in physical space, only the Second Cohort used the left-right axis of signing space to represent changes in perspective on an event (e.g., the narrator's perspective vs. a participant's perspective). From this, the authors suggest that devices using signing space to indicate changes in perspective emerged only after NSL established more basic conventions for left-right spatial contrasts in descriptions of physical space.

Thus, to reiterate, to understand phenomena of more general interest, like grammaticalization, phonological combinatoriality, all key properties of both syntax/semantics/pragmatics and phonology/phonetics, we may need to first understand conventionalization and reduction.

### **Theories of conventionalization and reduction**

The above discussion should make clear that conventionalization and reduction are (1) important phenomena that occur in language emergence, and (2) *occur together*. However, it is not clear how each phenomenon occurs, and even less clear what causal network(s) might be connecting them (i.e., whether one influences the other, they influence each other, and/or additional variable(s) affect both). We thus now turn our attention to considering theories of conventionalization and reduction, and theories of why they are related.

One possible theory of language emergence is that strong, central institutions dictate to its members how the language should emerge -- in our case, how words should be produced. But a role of social institutions, while plausible in the invention and enforcement of some conventions

and norms (for example, laws enacted by legislative bodies like the US Congress, or linguistic norms recommended by the French Academy), is not particularly plausible for truly de novo language emergence. Theories relying on the role of institutions do not explain whether, or how, it is possible for conventions to emerge when social institutions are not already in place to guide the process, as is commonplace in situations of language emergence (where the institutions only emerge *after* the users of the language have a means -- their language -- to organize).

Alternatively, language emergence -- and particularly conventionalization and reduction - - could be the emergent result of many, local interactions between individual language users attempting coordination. In other words, language may *self-organize*. In fact, this is the approach most scholars of language emergence take (Bybee, 2006; Galantucci, Garrod, & Roberts, 2012; Richie et al., 2014; Baronchelli, Gong, Puglisi, & Loreto, 2010; Sandler et al. 2011; i.a.).

Consider conventionalization. Most scholars posit that, even if a population does not share a form for a meaning initially, pairs or groups of individuals can interact and gradually adjust their future behavior based on such interactions, and thus gradually come to agree on a single form for a particular meaning. Beyond this, however, there is not much agreement about how conventionalization proceeds. There is recognition that conventionalization is influenced by social network structure, but even different experiments and modeling simulations all concerning lexical conventionalization find inconsistent effects (e.g., see Gong et al., 2012; Richie et al., 2014; Centola & Baronchelli, 2015; Hall et al., 2016 for examples). The lack of theoretical consensus on conventionalization is likely due to the proliferation of various experimental conditions and models. To illustrate, conventionalization in the experiment of Hall et al. (2016) and model of Richie et al. (2014) occurred over sequences of gestures, but over single, holistic

names in the models of Centola and Baronchelli (2015) and Gong et al. (2012).

One way to counteract such model proliferation is to strip the models down to a bare framework, and then add in features until required components of an effective communication system (e.g., conventionality) emerge. Indeed, this is the approach taken by Spike, Stadler, Kirby, and Smith (2016), who found that emergence of conventional, unambiguous lexicons required (1) transmission of referential information from speaker to hearer (that is, the hearer must have some information besides the utterance of what the speaker is referring to; for example: a speaker might point to an intended object; or perhaps only a subset of possible referents exists in the context of the utterance) (2) an anti-ambiguity bias (e.g., regularly removing homonyms from the lexicon), and (3) some form of limitation in the population's memory of past form-meaning pairings. However, as they point out, even within these constraints there are a great many possible model architectures that create effective lexicons, and different empirical settings could even engage different cognitive mechanisms. Thus, to further narrow down the field of models, we need to conduct detailed comparisons between model behavior and empirical data. To date, however, investigation of conventionalization (and language emergence generally, even) has typically not used computational models to account for qualitative (except in the grossest sense: does the model produce agreement in the end or not?) or quantitative patterns in empirical conventionalization, Richie et al. (2014) and Centola & Baronchelli (2015) being exceptions. In other words, computational models of conventionalization often go untested by empirical data. This is one gap to be addressed in this dissertation.

Consensus is also lacking among theories of reduction. A common factor among most or all theories of reduction is that frequent or otherwise predictable forms are those that usually

become reduced, but theories differ in *why* frequency or predictability have the reductive effects that they do. Listener-oriented theories (also known as common ground or audience design theories; Fowler and Housum, 1987; Galati & Brennan, 2010; Arnold, Kahn, & Pancani, 2012; Kahn and Arnold, 2012) claim that speakers lengthen or shorten utterances when they expect comprehension will be hard or easy, respectively, for the listener. Frequent words and words already in the common ground of an interaction are easy for the listener to predict/retrieve, and hence can be shortened by the speaker, saving their own time and effort, without sacrificing the listener's comprehension. Such accounts also predict that speakers will augment or reduce their utterances based on past success or failure in communicating with interlocutors, i.e., reduce utterances after success, and enhance them after failure, a prediction borne out empirically (at least with enhancement following failure, e.g., Buz, Tanenhaus, & Jaeger, 2016; Schertz, 2013; Stent, Huffman, & Brennan, 2008). Speaker-oriented accounts, by contrast, posit that reduction is simply due to facilitated (or primed) production that results from a word being produced frequently (e.g., Lam & Watson, 2014; Jacobs et al., 2015). Facilitated, primed, or practiced processes, whether they be making a sandwich or singing the ABC's, are able to be executed more quickly, and hence be reduced in duration or length.

At least two gaps emerge from reviewing this literature on reduction. First, theorists of reduction usually concede that these two families of theories are not mutually exclusive, but it still remains largely unknown which mechanism(s) cause(s) which of the many empirical cases of reduction, the cases of current interest being morphological reduction in language emergence settings. In particular, theorists of reduction like Jaeger and Buz (2016, pg. 17) find it “counter-intuitive to argue that the omission of optional elements (e.g., optional arguments, adjuncts, function words) is a consequence of these elements being easy to produce”. Second, work on

reduction has mostly been empirical and has been used to construct verbal, often vague, models - there are not, to this writer's knowledge, any computational models of reduction<sup>6</sup>. Thus, to address these gaps, we will attempt to develop a *computational* model of reduction in number of *discrete* elements. Further, the model we develop may not clearly map onto either speaker- or listener-oriented accounts.

### **Theories of the *relationship between* conventionalization and reduction**

Scholars have thus elaborated a number of theories of conventionalization and reduction independently. As pointed out in an earlier section, these two phenomena often co-occur in language emergence, but this correlation has received even less theoretical development. A number of possibilities have been suggested. A common one is that conventionalization leads to forms being produced frequently, which, as pointed out above, is in turn a cause of reduction (whether through automation/practice for the speaker or ease of retrieval for the listener). Bybee (1999, pg 223), for one, suggests that “[w]ords and phrases that are used more undergo...reduction as part of the move to automate speech: boundaries are obscured and segments and syllables may disappear into the mass of co-articulated gestures”. She thus seems to favor a speaker-internal account, but gets no more specific about the particular mechanisms involved. Namboodiripad et al. (2016, pg. 8) similarly very briefly suggest that reduction correlated with conventionalization results from “routinization of familiar articulatory targets” (which, again, are familiar because they are conventionalized).

There thus seems to be a tendency among language emergence specialists to think that conventionalization and reduction interact in the following way: conventionalization causes

---

<sup>6</sup> The closest to a computational model of reduction would be work like Pierrehumbert (2001)'s model on lenition (roughly, weakening) of consonants, as in the change of Latin ‘pater’ to English ‘father’ (/p/ and /t/ weaken to /f/ and /th/, respectively). Lenition is often understood as a result of articulatory effort reduction. But this does not pertain to the reduction in count of discrete elements occurring *over time*, which we are concerned with here.

frequent usage of particular forms, which causes facilitated/automated access of such forms *in the speaker*, which results in reduction. However, these suggestions have been somewhat vague (again, partly owing to them being purely verbal theories/models), and similarly, as noted at the end of the last section, it is not clear that this theory could account for *morphological* reduction (as opposed to phonetic reduction). Further, there are other possible ways in which these two processes could interact. For one, it is entirely possible that listener-oriented reduction processes mediate the observed correlation between conventionalization and reduction in lexicon emergence: conventionalization makes particular forms frequent, i.e. predictable and easily accessible by the listener, which allows the speaker to reduce them. A somewhat different listener-oriented link might rely on the findings mentioned in the last section that speakers modulate their utterances based on the perceived past success or failure of communication. That is, language users conventionalize, and then communicate more successfully in virtue of conventions, and then reduce more because of communication success. *This dissertation will pay special heed to investigating this theory experimentally.*

A final possibility – neither speaker- nor listener-oriented – is that conventionalization and reduction both emerge – or self-organize – as side-effects resulting from independently necessary components of language functions, like production, comprehension, and learning. In this theory, no agent is explicitly attempting to either (a) conventionalize with *all* other agents – just their interlocutor, or (b) adjust their lexicon to save effort in communication. Nor would automating production have a role in this theory. Such a theory possibly has intuitive appeal for its simplicity – it posits no structure beyond what is already necessary for communication and learning – but it is presently unclear in its precise operation. *The computational model of the present dissertation will clarify this theory.*

The foregoing discussion should make clear that there is still great uncertainty in how conventionalization and reduction proceed on their own, and even greater uncertainty regarding how they interact. Many a priori plausible theories exist. Part of the uncertainty stems from the fact that scholars usually don't look at conventionalization and reduction over time, but rather only at the beginning and end. This has made it difficult to examine the ways in which conventionalization and reduction influence each other. Therefore, it is important to investigate these two phenomena simultaneously in an empirical, longitudinal setting of lexicon emergence, and then use a computational model to understand what processes of language use and change could give rise to the empirical observations of conventionalization, reduction, and their relationship. This is the aim of the present dissertation. In particular, the empirical and modeling portions of this dissertation will evaluate the listener-oriented link discussed above – that communication success mediates the correlation between conventionalization and reduction – and develop the above-described alternative, self-organization-based theory and computational model of conventionalization and reduction.

## **Chapter 2 - Experimental Method**

We now turn to describing the experiment we carried out to investigate emergence of communication systems generally, and conventionalization and reduction specifically. The experiment we conduct falls into the field now known as Experimental Semiotics (Galantucci, Garrod, and Roberts, 2012), which is concerned with studying emergence of novel human communication systems in experimental settings. Such studies have been conducted in a variety of communication media – from gestures (Namboodiripad et al., 2016), to slide whistles (Verhoef et al., 2015), to drawing pads (Galantucci, 2005) – and have yielded systems possessing

many structures found in natural languages, like semantic compositionality (Kirby et al., 2008) and phonological combinatoriality (Verhoef, 2012; Verhoef et al., 2015). These studies have varied greatly in the ecological naturalism of their experimental design. Galantucci (2005) used a drawing pad because it was a novel medium which prevented the use of conventional symbols like letters and numbers. This enabled them to investigate communication system emergence independent of modality effects of speech or gesture. However, to the extent that most – possibly all – naturally emerging/emerged languages are spoken or signed, the modality effects of these mediums are of inherent interest to understanding how natural languages emerged. On the other hand, Namboodiripad et al. (2016) possessed a degree of ecological validity because it used gestures as its communication medium, but lacked in ecological validity because gestural systems only developed within *pairs*, while natural languages develop in *groups* of many people. For the present work, we attempt to capture a relatively high degree of ecological validity relative to other experimental semiotics tasks. In brief, our experimental method entails having groups of hearing, nonsigning adults play repeated games of “charades”, with the goal of these groups gradually developing shared – and reduced – referring expressions for the different objects in the task.

## **Participants**

We asked hearing undergraduates who had no experience with sign language to engage in a dyadic gestural communication task, in groups of four participants each, called “quads”. To complete the task, each quad needed to complete two sessions about one week apart. (This time delay was due to the logistical constraints of scheduling 4 participants and 2 experimenters, and does not play a substantive role in analyses.) Because partial data could not address the research questions, it was necessary to discard entire quads if the data were incomplete. This occurred due to one or more participants in a quad not showing up for or not completing one of the two

sessions (n=6 quads), experimenter or equipment error (n=1 quad), or having had previous experience with sign language or gesture, including having previously participated in similar gesture experiments (n=1 quad). In order to achieve a final dataset of 16 quads (64 individuals), it was necessary to enroll a total of 24 quads (96 individuals). The present data are composed of only 8 quads, which interacted in a fully-connected network, wherein each individual has opportunities to interact with every other individual (the other 8 quads interacted in a star network structure, in which only one participant interacts with the other three members of a quad, for a different study). All participants signed informed consent for the study, which was approved by the Institutional Review Board of the University of Connecticut. Participants received either course credit or \$10/session. Initially participants were paid a \$5 bonus for completing both sessions; this bonus was later removed when it became clear that it was neither necessary nor effective.

### **Design & Procedure**

Each participant was randomly assigned to a position within the quad, which we refer to as A, B, C, and D. Participants then paired off into rotating dyads as shown in Table 1. Participants took turns producing and comprehending gestured descriptions of 25 entities. Speech, writing, mouthing words, and audible sound effects were prohibited (although participants were allowed to silently mouth sound effects like MOO, as we found this was common in Nicaraguan homesign, Richie et al., 2014). Each participant had a booklet displaying a target image to describe, as well as an array of 25 images corresponding to the possible items that their interlocutor might describe. The 25 images were the same for both partners, but ordered differently. A low visual barrier prevented participants from looking at each other's booklets and arrays. Figure 1 shows the experimental setup.

After one participant described an item using gestures, the other participant would select the corresponding image from their own 5 x 5 array by silently pointing to one of the items. The experimenter recorded the selected item, and then said “ok” to indicate that the next trial could begin. Neither participant received any explicit, reliable feedback from the experimenter about the comprehender’s accuracy, but participants were free to signal uncertainty to each other using gestures. After the members of a dyad had described all 25 images to each other, they switched partners. The first “round” was completed once each participant had communicated with all assigned interlocutors. Participants completed rounds 1 and 2 during the first session, and completed rounds 3 and 4 approximately one week later, for a total of four interactions with the same three interlocutors.

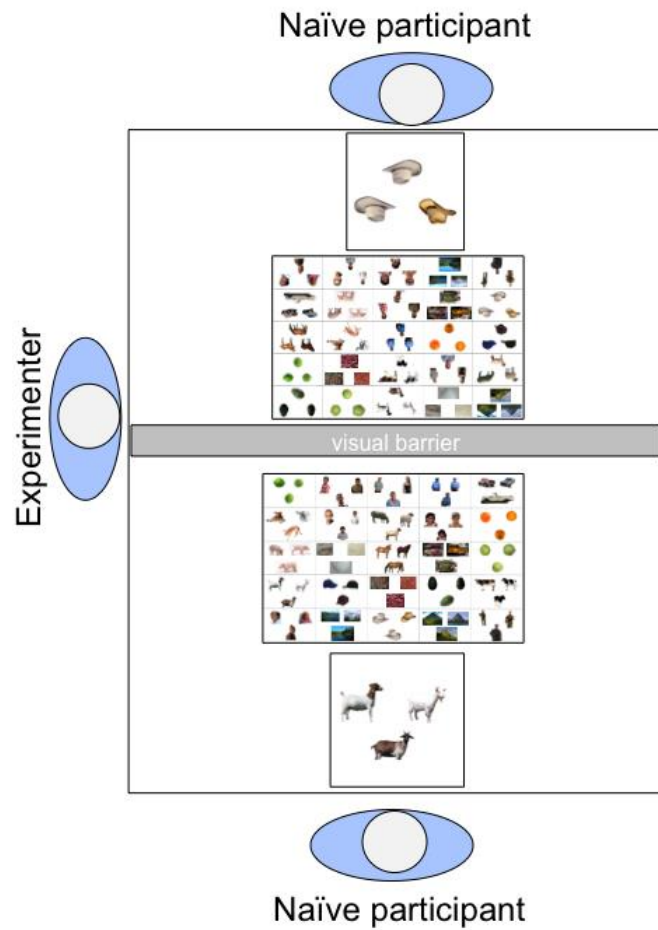


Figure 1. Experimental setup. One participant, the describer, sees their current target object on the booklet immediately in front of them. They utter a string of gestures to the comprehender across the table, who then selects an object from the array. Participants' arrays have the same objects ordered differently.

Day	Round	Total number of pairings so far (i.e., # of prior interactions among A, B, C, & D)	Room 1	Room 2
1	1	0	AB	CD
		2	AC	BD
		4	AD	BC
	2	6	AB	CD
		8	AC	BD
		10	AD	BC
2	3	12	AB	CD
		14	AC	BD
		16	AD	BC
	4	18	AB	CD
		20	AC	BD
		22	AD	BC

Table 1. Schedule of interactions.

## **Materials**

The stimuli were images of 25 items adapted from Richie and Yang (2013), listed in the Appendix. To encourage participants to create expressions that referred to types (e.g., “girl”) rather than tokens (e.g., a particular girl), each item was represented by 3 exemplars. The original Richie and Yang items were created to be recognizable to Nicaraguan participants; we selected a subset of those items that would also be familiar to American college students, and supplemented as necessary to ensure that all items had at least one semantically-similar competitor (e.g., ‘cow’ had the competitor ‘goat’).

## **Coding**

Participants all gave consent to be video-recorded as part of the study. These recordings were used for offline coding by the experimenters. For each of the 9600 gestured utterances obtained, an experimenter labeled each gesture according to the aspect of the target object that the gesture iconically represented. We will refer to these labels as ‘conceptual components’, or CC’s. This labeling was applied to all gestures that were produced between trial onset and when the interlocutor (correctly or incorrectly) selected an image from the target array. For example, if two participants produced a gesture for stroking a beard, but the handshape or motion is slightly different, we still code it as BEARD. The motivations for this coding scheme, which focuses on conceptual/iconic aspects of a gesture rather than fine phonetic/phonological aspects, were both practical and theoretical. Practical motivations included the fact that it was easier to code conceptual/iconic aspects rather than phonetic/phonological aspects. This scheme was also theoretically motivated because no reliable, corrective feedback was given to participants, and yet they communicated successfully, meaning that participants had to rely on iconicity to communicate, at least initially.

To gauge inter-rater reliability, a second experimenter independently coded 25% of the utterances. To measure inter-rater reliability, we chose Cohen’s Kappa. It is thought to be a more robust measure of inter-rater reliability than simple percentage agreement, as Cohen’s Kappa takes into account the amount of agreement expected due to chance alone. Cohen’s Kappa is defined as:

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$

where  $p_o$  is the percentage of agreement among raters, and  $p_e$  is the percentage of agreement expected due to chance alone (computed from the observed frequencies of the labels). We computed Cohen’s kappa for each combination of item and quad (e.g., the first full network’s ‘avocado’ utterances). For the broader dataset (the fully-connected networks analyzed here plus star networks) from which the present data were drawn, average Cohen’s kappa was .59, indicating moderate, bordering on substantial, agreement, according to Landis and Koch’s (1977) guidelines.

### **Measuring string similarity**

After coding, each gestured utterance was now represented by a set of conceptual components (e.g. MILKING, HORNS, DRINK, MOO, 4-LEGS, etc.). We quantify conventionalization within a quad with the Jaccard index, a measure of similarity among sample sets, defined as the size of the intersection of the sets divided by the size of the union of the sets. For example, the sets {HORNS, CHEWING, BEARD}, {HORNS, EARS, BEARD}, {HORNS, EARS}, and {HORNS, CHEWING} have a Jaccard index of 1/4 (0.25), because their intersection contains 1 element {HORNS} and their union contains 4 elements {HORNS,

CHEWING, BEARD, EARS}. Thus, a Jaccard index of 0 reflects total absence of conventionalization, while 1 reflects perfect conventionalization. We chose this measure for three reasons. First, it computes similarity simultaneously among more than 2 sets. This was important as we wanted to compute similarity among four participants simultaneously. Second, it is defined over sets of varying length; this was important as participants' utterances varied in length. Third, it is insensitive to order; this was important because participants varied the order of their gestures, even when using the same gestures.

## **Chapter 3 - Experimental Results**

This section is organized as follows: We first report qualitative impressions of the data (Tables 2a and 2b) that illustrate apparent similarity between our experimental data and naturalistic data from emerging sign systems like ABSL (Al-Sayyid Bedouin Sign Language). We then describe a set of exploratory analyses suggesting that participants conventionalize somewhat *arbitrary* systems (Hockett, 1960; Rescorla, 2015; figures 3 and 4), in a self-organizing (as opposed to dictatorial) manner (figure 5). We then report analyses showing that conventionalization and reduction are correlated across the entire dataset (figure 6), and for individual items (figures 7 and 8). Finally, we report two analyses testing a listener-oriented/communicative account of the link between conventionalization and reduction. In brief, these analyses find that participants reduce at higher rates after communication failure, a result that is inconsistent with the listener-oriented theory.

### **Experimental data have ecological validity**

Tables 2a and 2b show utterances drawn from our experiment, and from signers of ABSL (Meir et al. 2010), respectively. Though Tables 2a and 2b show responses for different objects (we didn't include 'stove' in our experiment), they show that in both the experimental and

naturalistic data, signers and gesturers typically produce multiple iconic forms for each utterance, often encoding similar information (e.g., objects' size and shape, their typical affordances, their typical actions). In addition, Morford and Kegl (2000) and Richie, Coppola, and Yang (2014) report similar utterances encoding objects' size, shape, affordances, and actions, for other basic objects in both Nicaraguan homesign systems and Nicaraguan Sign Language. The same appears to be true for homesign systems on Amami Island near Japan (Osugi, Supalla, & Webb, 1999). The apparent similarity between our experimental data and naturalistic data from a variety of cultures and types of systems (e.g., homesign, emerging sign languages) suggest a degree of ecological validity for our experimental data (in contrast, possibly, to the communication mediums of other experimental semiotics tasks, e.g., Galantucci, 2005).

<b>Gesture_1</b>	<b>Gesture_2</b>	<b>Gesture_3</b>	<b>Gesture_4</b>	<b>Gesture_5</b>	<b>Gesture_6</b>
small-round	eat	sour	wedge	small-round	peel
pick	small-round	eat			
small-round	peel	<sup>7</sup> not-large-round			
not-large-round	small-round	eat			

Table 2a. Sample of utterances for 'lime' produced by a participant in our experiment. Each row is a single utterance.

<b>Sign_1</b>	<b>Sign_2</b>	<b>Sign_3</b>	<b>Sign_4</b>	<b>Sign_5</b>
Turn	cook	wide-object		

---

<sup>7</sup> The gesture for 'not' – variously a headshake and/or arms crossed and then uncrossed repeated – is likely derived from conventional gestures among American English speakers meaning similar things, like 'no' or 'stop'. While we have not completely prevented participants from using these communication conventions, such usages are rare in our data.

Turn	fire	four	burner	fire
Turn	wide-object			
Cook	insert			

Table 2b. Sample of utterances for ‘stove’ produced by signers of ABSL (Meir et al., 2010). Each row is an utterance. We label the utterances in table 2a as ‘gestures’ and the utterances in table 2b as ‘signs’ as ABSL has shown myriad forms of linguistic complexity that our participants’ gestures have not.

### **Different quads conventionalize different systems, with some arbitrariness, in a self-organizing manner**

Figure 2 (blue curve) shows how the mean Jaccard index increased over time (as measured by “number of pairings so far”, see Table 1). We wanted to assess the extent to which this increasing jaccard index represented conventionalization, as opposed to mere regularization. That is, an increasing jaccard index could, a priori, simply reflect all participants, regardless of quad, discovering the single, optimal solution to communicating in the task. In contrast, groups arriving at *true* conventions would require a degree of arbitrariness of the conventions (Rescorla, 2015; Hockett, 1960). That is, different groups ought to arrive at different conventions. To assess this, we conducted two different analyses. In the first, a Monte Carlo simulation, we repeatedly randomly reassigned participants in our existing dataset to novel, shuffled quads, and recomputed the jaccard index for these shuffled quads. Figure 2 shows that these shuffled quads conventionalized much less than the real quads, and that shuffled quads did not conventionalize in any appreciable way, suggesting that members of a quad are converging with each other, rather than with the participants of the entire experiment. Nor are quads inevitably and “independently” arriving at an identical, optimal solution. The second analysis is displayed in figure 3. Here, we computed pairwise Jaccard indexes for both within-quad communication partners (right panel of figure 3), and between-quad communication partners (left panel of figure 3). As in the first analysis, the within-quad jaccard indexes are much higher than the between-

quad jaccard indexes, the former reaching an average of .75 and the latter reaching an average of .25 by the end of the experiment. Finally, according to Lewis (1969), for a regularity  $R$  to be considered ‘arbitrary’, there must be some plausible alternative regularity  $R'$  that could serve  $R$ ’s function. The English word ‘cat’ thus constitutes a convention, because there is the plausible alternative ‘gato’. We suggest that, here, for  $R'$  to be considered a *plausible* alternative that *could* serve  $R$ ’s function, a gestural system  $R'$  should accomplish the task of communication comparably well to  $R$ . We already demonstrated that different quads attain different systems  $R$ ,  $R'$ , etc.; but this does not demonstrate that these different systems suit the task of communication equally well. Figure 4, however, shows that they do. Figure 4 shows that error in the task declines over time *and varies little across the quads*. Thus, we argue that the different gestural systems that our quads converge on are properly *arbitrary*.

Finally, the emergence of language, of which conventionalization is a part, is often thought to be a *self-organizing* process (Barr, 2004; Gong et al., 2012; Richie et al., 2014; Sandler et al. 2011). Self-organization is a term that has been used to describe a very wide range of phenomena. In computer science and artificial life, self-organization arises in cellular automata like Conway’s Game of Life, where simple beings die or spring to life based on their neighbors death/life, creating complex structures above the level of individual organisms (Gardner, 1970). In agent-based modeling (the kind of modeling often applied to language emergence, and which we will carry out in the present work), the term is used to describe systems of interacting agents that exhibit systematic behaviors. For example, Reynolds (1987) showed that “coordinated” flocking behavior emerges (“self-organizes”) among agents following very simple rules (e.g., maintain a certain distance from neighbors and maintain a similar heading). Kukona, Cho, Magnuson, and Tabor (2014), provide a definition of self-

organization that will suffice for us:

Self-organization refers to situations in which many, small, autonomously acting but continuously interacting elements exhibit, via convergence under feedback, organized structure at the scale of the group.

Such a definition could, in principle, apply to our current experiment: individual, autonomous interlocutors interact with one another, and through feedback, converge on conventions shared across the whole group. However, it is worth more stringently probing the activity of our participants for evidence of self-organization. In the present data, self-organization would mean that no single member(s) of the language community lead or dominate, and instead, members create co-equally<sup>8</sup>. This self-organizing account contrasts most starkly with the dictatorial or ‘institutional’ account described in the introduction, where single actors – whether individuals or institutions – dictate the convention to be followed. To assess, to a first approximation, whether self-organization reasonably characterized the quads in our study, we computed how much each participant changed their utterances – whether reducing or adding gestures – for a given item from interaction to interaction. If conventionalization were self-organizing in the specific sense defined above, then we would expect different participants within a quad to change their utterances to a similar degree. Figure 5 shows that, for each quad, this is indeed the case. Each of the 32 participants, represented by a different curve in a different subplot, changes roughly as much as their quad-mates, suggesting that no one participant within a quad controls the evolution of the quad’s system. If quads evolved in a dictatorial fashion, on the other hand, then each quad would have one participant who changed little over time, represented by a curve close to  $y=0$ .

---

<sup>8</sup> At least, members *within* a generational cohort create more or less co-equally. *Different* generational cohorts seem to have different effects on language emergence (e.g., Senghas & Coppola, 2001).

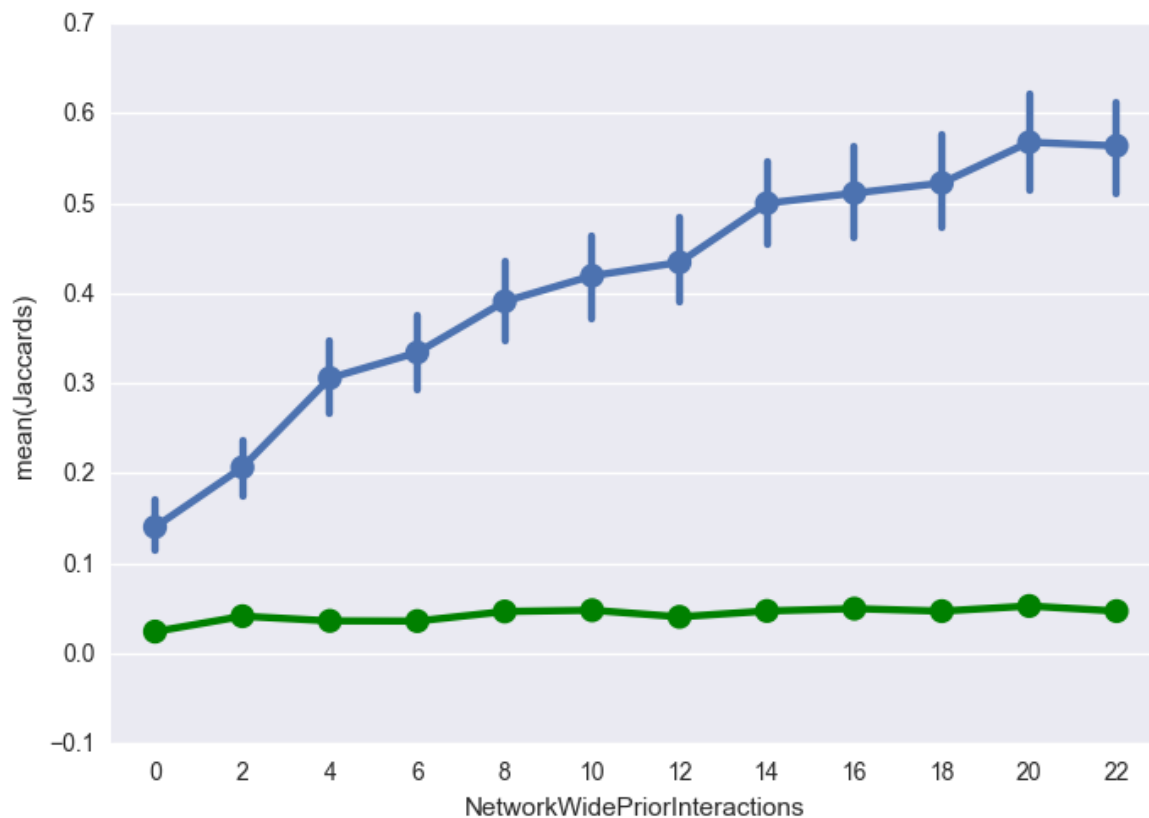


Figure 2. On the x-axis is time, as measured by pairings in the quad so far. On the y-axis is mean jaccard index, the measure of conventionalization. The blue curve represents the jaccard index within real quads. The green curve represents the jaccard index within shuffled quads. Real quads converged on shared systems substantially, while simulated quads do not. This finding, coupled with the finding that different quads perform roughly equally at the task, suggests that quads are creating proper conventions, in the sense that quads choose different but equally effective alternative systems.

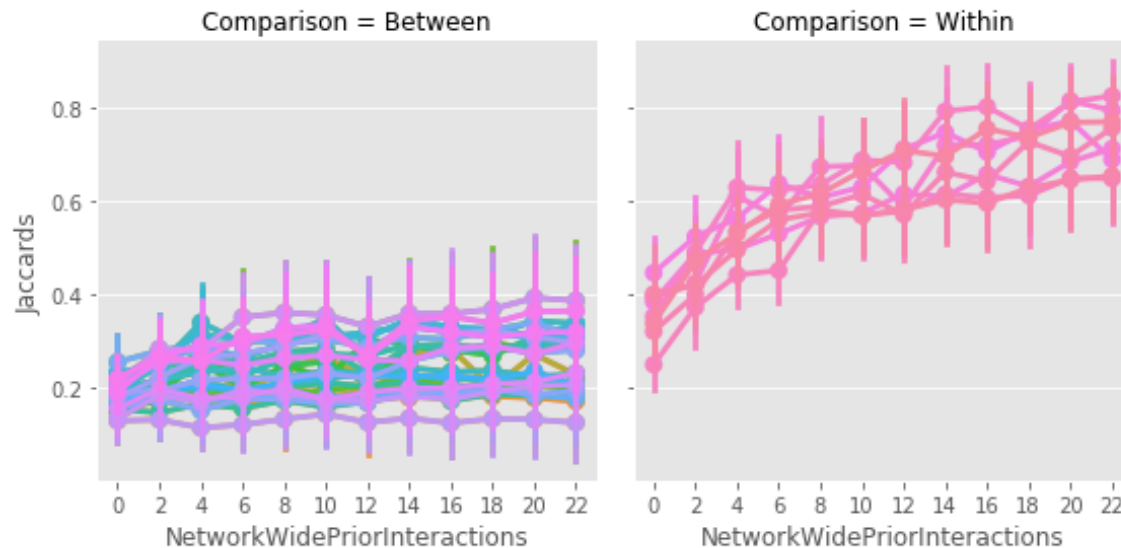


Figure 3. This plot demonstrates that quads attain idiosyncratic, quad-specific agreement. Each quad is represented by a single line. The pairwise jaccard index is similar to the 4-way jaccard index used elsewhere, but is a more liberal measure of conventionalization, allowing for merely local conventionalization. The plot on the right demonstrates robust within-quad conventionalization, and was generated as follows: at each time point and item, we computed the average jaccard index between all possible pairs of *within-quad productions* (4 choose 2 = 6 pairs of productions for each combination of quad, item, and time). The plot on the left demonstrates the low between-group conventionalization, and was generated as follows: Each curve represents a comparison between one group and a different group (there are 28 such possible comparisons). For each time point and item, we computed the average jaccard index between all possible *between-group* pairs of productions, i.e., between F01-A and F02-A, F01-A and F02-B, and so forth. Color represents different between-quad (left-panel) or within-quad (right-panel) comparisons, but we omit a legend indicating the mapping from color to comparison for space concerns, and because the different comparisons are not semantically meaningful as, say, item or time point would be. Within-group conventionalization is higher even at  $t=0$  because each participant produces half of the utterances at  $t=0$  *after* having seen their partner’s production for the same item. So,  $t=0$  already enables some quad-specific conventionalization.

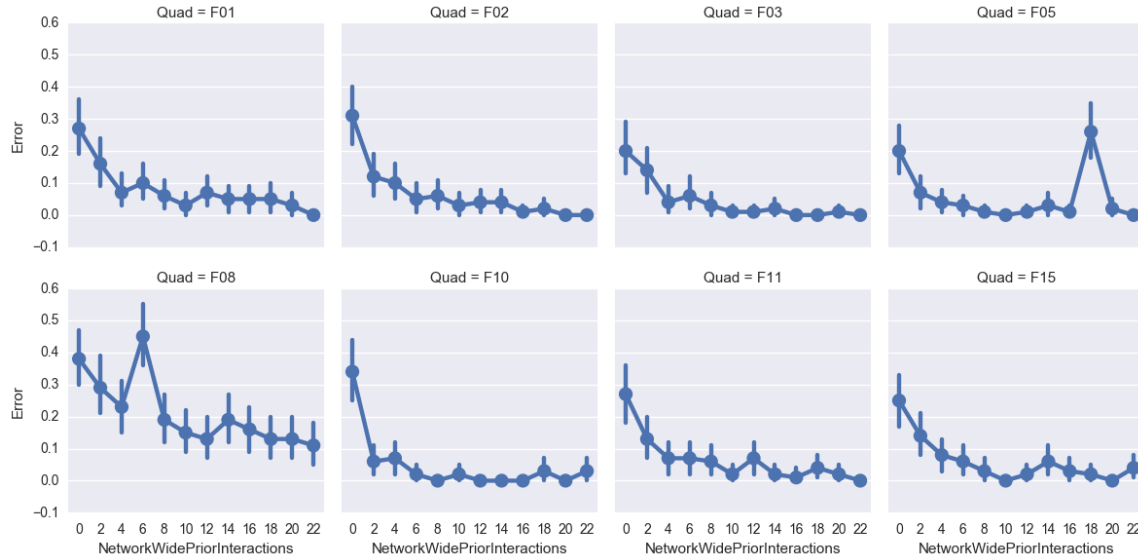


Figure 4. X-axis is time. On the y-axis is the rate of error. Error decreases across the experiment fairly consistently, and fairly comparably across quads, providing evidence that all quads reached equally optimal systems.

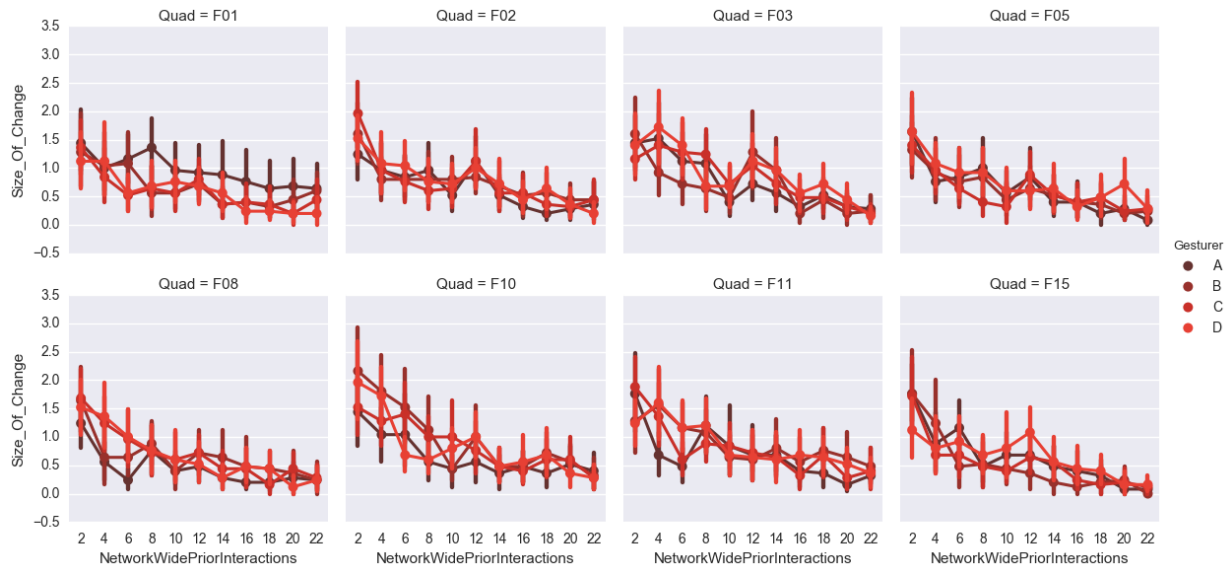


Figure 5. Each subplot represents a single quad. Each hue represents a different member of a quad. On the x-axis is time. On the y-axis is the magnitude of the difference between consecutive utterances for the same object within a given individual. For example, if a participant produced {HORNS, MILKING} and then {HORNS, MOO} for 'cow', the magnitude of their change would be 2 (1 reduction plus 1 addition). It is not important to distinguish different hues. In fact, that participants' lines mostly overlap suggests that the degree of change is similar across individuals within a quad, suggesting that no one participant in a quad dictates the process.

Rather, participants co-create, or self-organize.

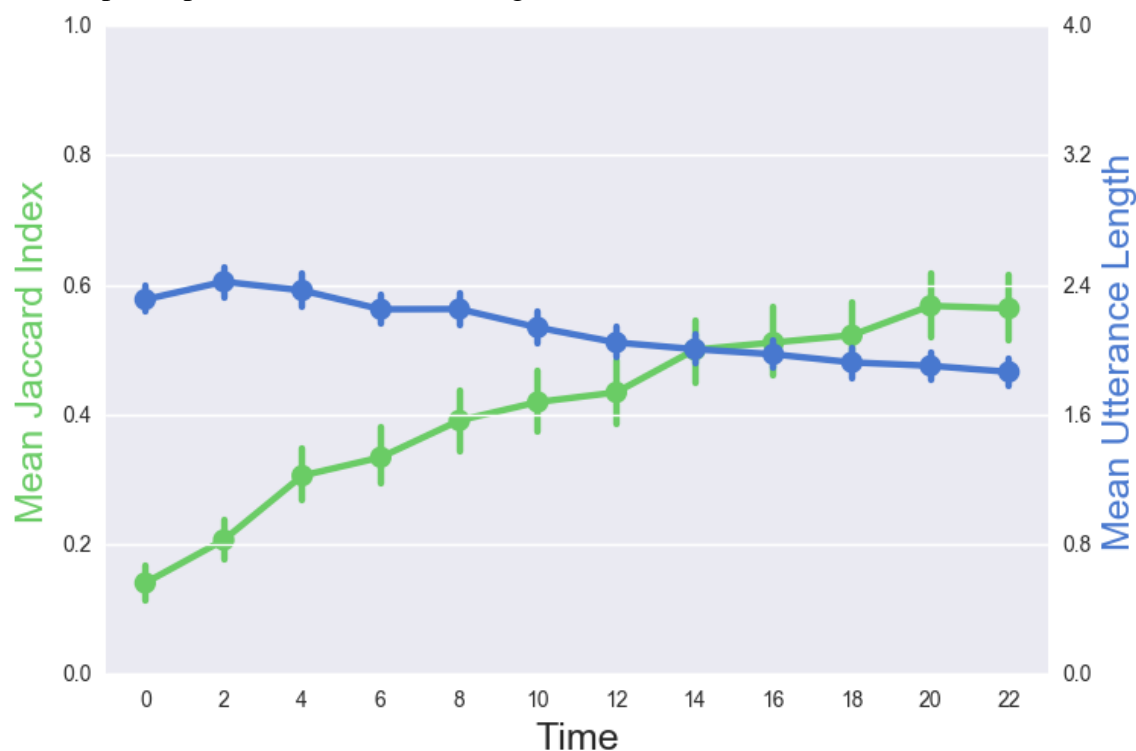


Figure 6. On the x-axis is time. In green is the mean jaccard index, and in blue is the mean utterance length, as measured by the number of gestures in an utterance. Conventionalization and reduction are correlated: while quads conventionalize, they reduce.

### Conventionalization and reduction are correlated

The previous section demonstrated conventionalization in our experiment. We now turn to showing that *reduction* accompanies conventionalization in our experiment. Figure 6 shows conventionalization and reduction over time, averaged across all quads and items (and participants, in the case of reduction); as quads conventionalized, they also reduced. We also investigated this correlation across items (see figure 7). For each item, we conducted ordinary least squares regressions predicting jaccard index from time (pairings in the quad so far), and utterance length from time. We then extracted these slopes, and conducted a Pearson's correlation between jaccard slopes and utterance length slopes. Jaccard slope was negatively correlated with utterance length slope, such that those items that conventionalized more drastically (increased jaccard index) also reduced more drastically (decreased utterance length),

$r(23) = -0.57, p = .003$ . See Figure 8 for a scatterplot of jaccard slopes against utterance length slopes.

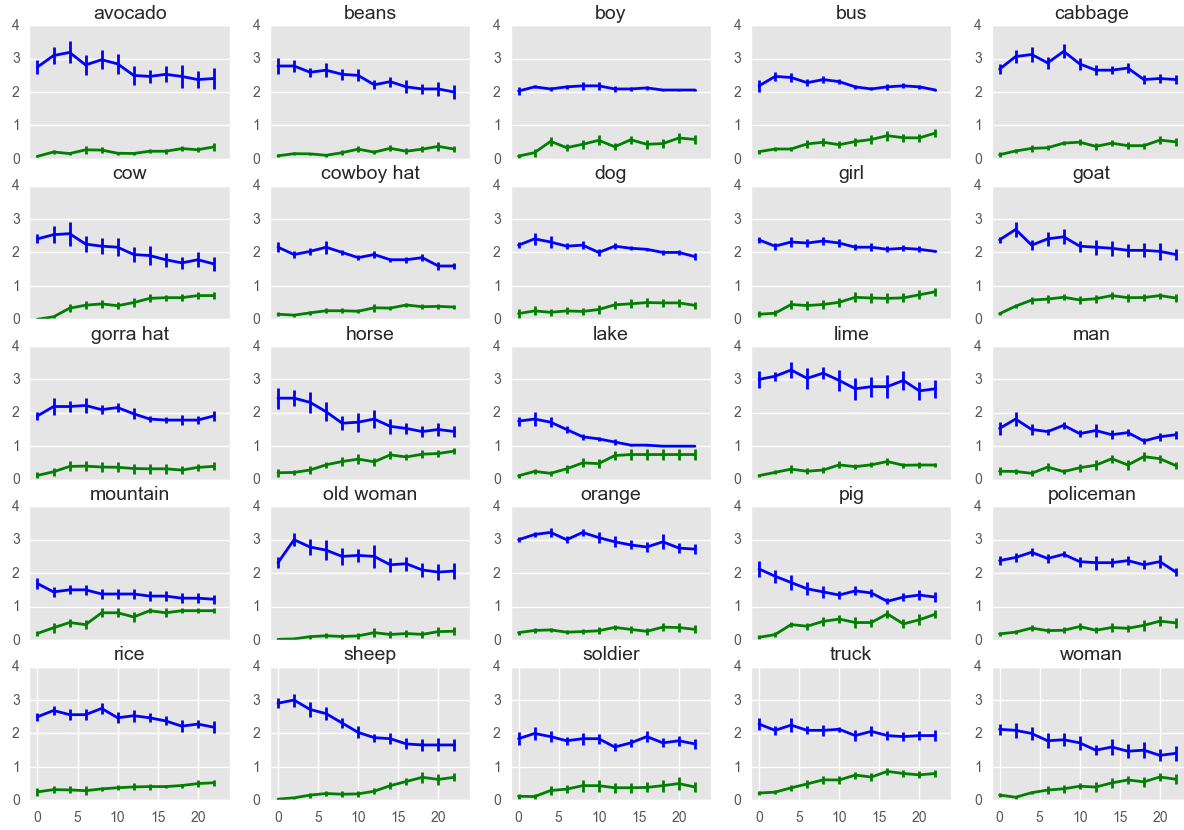


Figure 7. Conventionalization vs. reduction for each stimulus item. In each subplot, the x-axis represents time (# of prior interactions) in the experiment. The blue curves represent degree of reduction (measured by response length, the number of discrete gestures in a response), while the green curves represent degree of conventionalization (measured by Jaccard Index). It somewhat appears that the items that conventionalize more (increase most in Jaccard Index) are also those that reduce more (in response length). See text and figure 8 for following up on this possible relationship.

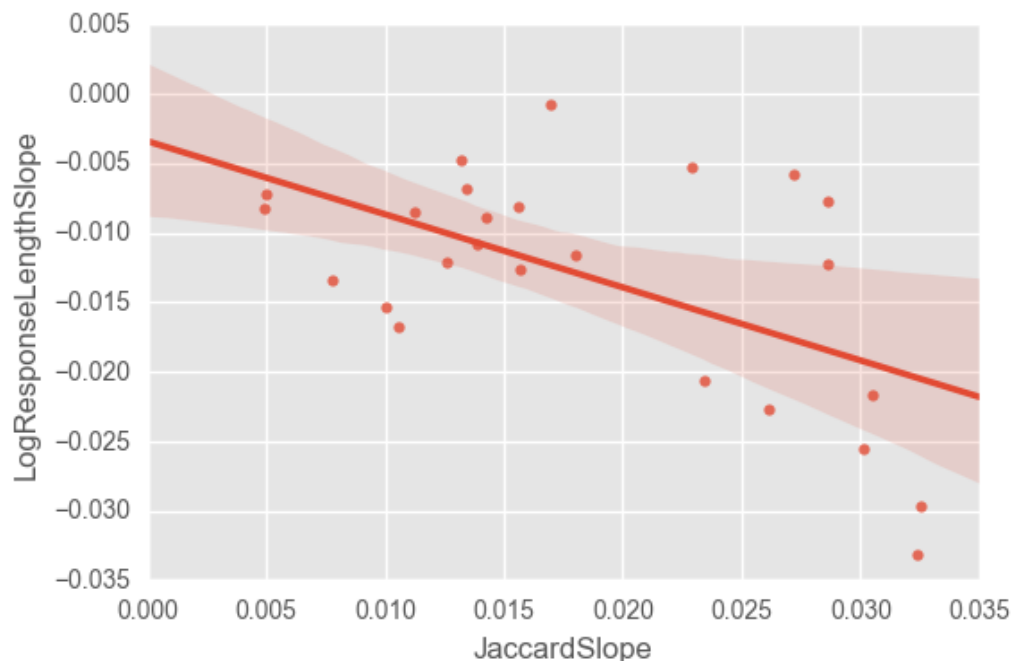


Figure 8. Slopes of jaccards and (log) response lengths, and a linear regression line. Those items that conventionalize more (higher Jaccard slope), also reduce more (more negative response length slope).

### Analyzing the relationship between reduction and communication success

Recall that under various forms of listener-oriented, audience design or communicative accounts of reduction (Brennan & Hanna, 2009; Buz, Tanenhaus, & Jaeger, 2016; Jaeger & Buz, 2016; i.a.), a language producer is expected to reduce an utterance for a particular message following successful communication of that message (or, conversely, enhance a signal following failed communication). Thus, a plausible account of the relationship between conventionalization and reduction is that conventionalization causes communication success, which in turn causes reduction. In this section, we report an analysis testing whether participants do in fact reduce utterances for a given object after successful communication on the previous trial the object was described.

We investigated this question with two analyses. We first conducted a mixed logit model (Jaeger, 2008) predicting whether a participant's utterance for a given object is reduced or not,

depending on the success or failure of the participant's previous utterance for that object. Previous success/failure was thus our single fixed effect. We also accounted for several random effects: item, quad, *standardized* time (i.e., centering and scaling our network-wide prior interactions variable), gesturer, comprehender, quad\*gesturer interaction, and the quad\*comprehender interaction), maximizing the random effect structure justified by the data (i.e., including random effects for slopes and intercepts; Barr, Levy, Scheepers, and Tily, 2013). Contrary to predictions, we found that reduction was correlated with previous *failure*,  $b = -0.743$ ,  $p = 1.97 * 10^{-5}$ . This effect was robust to different specifications of fixed and random effects structure (e.g., including time, gesturer, and comprehender as fixed effects). Our second analysis revealed the same finding: we extracted odds-ratios (of the odds of reduction following failure relative to the odds of reduction following success) for the different quads, different items, and different participants, and asked if the number of odds-ratios reflecting greater rates of reduction following success compared to failure were greater than could be expected due to chance. Indeed, all 8 of 8 quads had greater odds of reducing following failure ( $p=.008$ ), as did 21 of 25 items ( $p=.0009$ ), and 27 out of 32 participants ( $p=.0001$ ).

These two analyses suggest the following: participants did *not* reduce at higher rates more after communication success as would be expected under certain listener-oriented accounts, but rather did the opposite, reducing at higher rates after communication failure. This unexpected finding thus suggests developing a different theory of conventionalization and reduction, implemented in a computational model.

## Chapter 4 - Computational Modeling

We now turn our attention to modeling the phenomena of conventionalization and reduction, particularly as they manifest in the experiment just described. We proceed by (1) highlighting phenomena (again, particularly features of the experimental setup and results) that we attempt to model, (2) describing a previous model of conventionalization (Richie et al., 2014) and how it does not model these phenomena, (3) describing how we modify that model to capture (1), and (4) reporting behavior of this model.

To foreshadow the results of this section, our model demonstrates that conventionalization, reduction, and communication success – and particularly the relationships among them -- can be explained by assuming iconicity in the gestural modality, and simple probabilistic mechanisms of communication and learning. Further, the correlation observed in the lab and in the wild between conventionalization and reduction requires no links of speaker-oriented or listener-oriented varieties. Rather, both conventionalization and reduction can result from agents (human or simulated) limiting their associations between gestures and objects to a mutually agreed upon narrower subset of gestures for a given object. This narrowing makes agents more certain of what to say when attempting to communicate an object, which translates into more concise expressions.

### **Key aspects of the experiment to be modeled**

Several desiderata for a model emerge from our experiment. We describe these now:

1. **Participants produce iconic gestures.** Further, certain gestures only appeared with certain objects. For example, MILKING appears with ‘cow’ and ‘goat’, but not ‘truck’. Since participants come to the task with similar experience and conceptual representations for each object, they use similar iconic gestures, at least initially, and objects already differ substantially in the iconic gestures associated with them.

2. **The comprehender in an interaction must *choose* an object from among a set of possible objects** *given* the producer's utterance. For communication to regularly succeed, this will require that participants use different gestures for different objects.
3. In the experiment, producer and comprehender receive **no explicit, reliable information about communication success or failure** (this fact does not jeopardize communication and learning because of the iconic gestures – somewhat unique to each object – described above).<sup>9</sup>
4. After participants within a quad interact, **different quads conventionalize different systems** for the same objects (see figures 2 and 3).
5. Participants simultaneously **conventionalize** referring expressions, **reduce** those expressions, *and* increase their **success** at the task (figures 4 and 6 and section *Conventionalization and reduction are correlated*).
6. As in naturalistic communication, there seems to be a link between communication success and conventionalization. That is, to be understood, interlocutors must *share* a communication system to some extent. However, in our experiment, we found that **while conventionalization as measured by the jaccard index was initially quite low** (blue curve in figure 2), **communication success was nevertheless quite high** (figure 4). (This high initial success despite low conventions is enabled by the iconicity referred to above.)

---

<sup>9</sup> Interlocutors were anecdotally observed to sometimes signal uncertainty with facial expressions, hesitation in gesturing or picking a target, and so forth. However, we left uncertainty signaling out as a feature to be modeled, because (1) it was not a feature of interest to the original coding scheme for conventionalization, and (2) it is unclear how much effect it could have on participants' development of gestural systems. Future work, however, could recode our video data for such signals of uncertainty, to assess the extent of their impact.

7. Finally, contrary to predictions of audience design/listener-oriented/communication ease theories, **participants in our experiment reduced *more often* after communication failure than after communication success.**

We will refer back to these desiderata when describing our new model, and the behavior it produces. We will show that the model satisfies all seven desiderata.

### **A previous model of conventionalization of referring expressions in an emerging signed system**

In the introduction we referred to a model of conventionalization of gestural referring expressions, devised by Richie, Yang, and Coppola (2014). This model was built with the aim of understanding effects of social network structure on conventionalization in cases of naturalistic gestural/signed language emergence (which was also the question of primary concern in Hall et al., 2016, which generated the experimental data for the present dissertation), cases which, as discussed earlier, bear striking resemblance to the present empirical data. Thus, this model provides a reasonable starting point for the present modeling aims. The architecture of the model is quite simple. An agent's lexicon is just a matrix of probabilities for producing particular gestures (the columns in the matrix) for particular concepts/objects (the rows), with each probability randomly initially set (i.e., before agents interact) between 0 and 1. Uttering a string for an object is merely sampling the probabilities of gestures for that object. For example, if an agent's vector of probabilities for an object is [.9, .9, .1, .9], they will likely produce a string of [1, 1, 0, 1], meaning the first, second, and fourth gestures are present. A listener likewise samples their probabilities for that object, and communication succeeds with a probability depending on how different (by Hamming distance) the speaker and listener utterances were (an 'analysis by synthesis' approach). If communication succeeds, the listener rewards or punishes their probabilities for the gestures associated with that object such that they are more likely to produce

strings like those the speaker produced. More specifically, they reward probabilities for present gestures and punish probabilities for absent gestures according to a reinforcement learning scheme (Bush & Mosteller, 1951). Over time, agents interacting in this way come to possess shared lexicons where they are likely to produce similar strings for the same objects.

The Richie et al. (2014) model does not account for several features described in *Key aspects of the experiment to be modeled*. First, because the model independently and identically distributed each gesture-object probability between 0 and 1, there was no modeling of the iconicity space described in the first paragraph of *Key aspects*. Initial distributions for an object were overwhelmingly likely to overlap considerably with distributions for other objects. Second, the Richie et al. model stipulated that comprehenders already knew the producer’s intended object (for example, this would be the case if the producer pointed to the intended object in the scene). Thus comprehenders were not required to infer the target object, as they were in the experiment and presumably in (many or most) natural settings. Third, communication success depends entirely on conventionalization in the Richie et al. model, while these two factors dissociate somewhat in the current experiment. Finally, and most importantly, there was no tendency toward reduction in the Richie et al. model: agents were equally likely to conventionalize on a probability vector of [1,1,1,1], which would reliably produce utterances 4 gestures long, as they were to conventionalize on a vector of [1,0,0,0], which would produce utterances 1 gesture long.<sup>10</sup>

	<b>Richie et al. (2014)</b>	<b>The present model</b>
<b>Initial lexicon</b>	Each gesture is an	Overlapping distributions

<sup>10</sup> In fact, agents could conventionalize a vector of all 0 probabilities, which would produce empty utterances. This outcome is less and less likely the more gesture slots agents must conventionalize over (i.e., the more columns in the lexicon matrix).

	independently, identically distributed probability	representing semantically related objects
<b>Speaking</b>	For the target object, sample each gesture’s probability	Sample target object’s distribution for gestures until target recognized
<b>Listening</b>	Analysis by synthesis	Infer target
<b>Update</b>	Reinforcement learning: reward present gestures, punish absent ones	Move guessed object’s distribution toward utterance’s distribution

Table 3. A comparison between the architectures of Richie et al. (2014) and the present model.

### A new model of conventionalization and reduction

To address these mismatches with the considerations raised in *Key aspects*, we devise a new model that bears some resemblances to Richie et al. (2014; see table 3 for comparison between the Richie et al. model and the present model). Most importantly, agents’ lexicons are still probabilistic mappings from objects to gestures. However, in the new model, agents’ vectors of probabilities for a given object are now probability *distributions*. That is, while in the original model a vector like [.9, .9, .1, .9] was possible, now the individual probabilities must sum to 1, so an allowable vector would be something like [.33, .33, .01, .33]. As will be explained soon, this change enabled modeling of reduction. To model the iconicity phenomena described so far (desideratum 1), agents’ lexicons start with uniform probability distributions over identical subsets of the gestures, with different objects’ subsets overlapping to a parameterized degree. Table 4a illustrates this setup.

Further, a speaking agent now produces an utterance by sampling their distribution (with replacement) for the target object until they recognize (guess) the target object from their own utterance. This speaking process can be understood as the agent communicating about the object until the utterance clearly specifies the object to them. The listener then guesses the target object using the same comprehension process (desideratum 2). Finally, listeners, *simply assuming they guessed the target correctly*, adjust their probability distribution for the guessed object by aligning it with the distribution of gestures in the utterance. At no point do speaker or listener

exchange information about each other's success or failure, or the actual intended target

(desideratum 3). In more detail, the process works as follows:

1. The speaker samples a gesture from their distribution for the target object, with replacement. This utterance constitutes a distribution over gestures.
2. For every object distribution  $x$  in the speaker's lexicon:
  - a. Compute Kullback-Leibler Divergence between the utterance distribution and the object distribution  $x$ . KL-Divergence for discrete distributions is defined as:

$$D_{\text{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

Where  $i$  is an index for every outcome (here, gesture).

- b. Squash the KL-Divergence to the interval [0,1] via:

$$\frac{1}{1 + e^{-kl\_divergence(speaker\_string, object\_probs)}} * 2 + 1$$

- c. Convert the squashed KL-Divergence to a similarity score between 0 and 1 via:

$$1 - \text{squashed\_kl\_divergence}(\text{speaker\_string}, \text{object\_probs})$$

3. Normalize the set of utterance-internal\_object similarities to a probability distribution over objects by dividing every similarity by the sum of the similarities:

$$\text{Distribution over objects} = \text{similarities} / \text{sum}(\text{similarities})$$

4. To guess an object, sample the distribution once.
5. If the target has been guessed, the utterance is finished. Otherwise go to step 1, and sample and add another gesture to the utterance. (In other words, the speaker repeats steps 1 through 5 until the speaker guesses the target.)
6. The *listener* now guesses the target using their own lexicon, and steps 2-4 (i.e., they only guess the target once).
7. The listener updates their distribution for the guessed object  $m$  as follows:
  - a. Let  $Y$  = the utterance's vector of probabilities for each gesture
  - b. Let  $Z$  = the listener's vector of probabilities for object  $m$
  - c.  $\text{delta} = \text{learning\_rate} * (Y - Z)$
  - d.  $Z' = Z + \text{delta}$
  - e. Renormalize the new probabilities  $Z'$  so they sum to 1:
    - i.  $Z' = Z' / \text{sum}(Z')$

See the next page for a simple example of this process in action.

	cow	goat
HORNS	.33	.33
MILKING	.33	.33
BEARD	0	.33
DRINK	.33	0

	cow	goat
HORNS	.35	.33
MILKING	.30	.33
BEARD	0	.33
DRINK	.35	0

Table 5. On the left is the initial lexicon for both agents A and B. On the right is the updated lexicon agent B obtains after listening to agent A. The current target object is ‘cow’.

1. Speaker samples ‘cow’ distribution, utters **HORNS**. Utterance is distribution **[1, 0, 0, 0]**.
2. For every object distribution  $x$  in the speaker’s lexicon, compute the similarity between the utterance distribution, and the object distribution.
  - a. ‘cow’  $\Rightarrow 1 - \text{squashed\_kl\_divergence}([1, 0, 0, 0], [.33, .33, 0, .33]) = .67$
  - b. ‘goat’  $\Rightarrow 1 - \text{squashed\_kl\_divergence}([1, 0, 0, 0], [.33, .33, .33, 0]) = .67$
3. Normalize the set of utterance-internal\_object similarities to a probability distribution over objects by dividing every similarity by the sum of the similarities:
  - a. ‘cow’  $\Rightarrow .67 / (.67 + .67) = .5$
  - b. ‘goat’  $\Rightarrow .67 / (.67 + .67) = .5$
4. To guess an object, sample the distribution once.
5. If the target has been guessed, the utterance is finished. Otherwise go to step 1, and sample and add another gesture to the utterance.
  - a. **For the purposes of illustration, suppose the speaker incorrectly guesses ‘goat’. The speaker now continues uttering gestures.**
  - b. Speaker samples ‘cow’ distribution, utters **DRINK**. Utterance is distribution **[.5, 0, 0, .5]**.
  - c. Recomputing steps 2 & 3 yields new probabilities for the objects: ‘cow’=1, ‘goat’=0. **The speaker correctly guesses ‘cow’ and stops uttering gestures.**
6. The *listener* now guesses the target using their own lexicon, and steps 2-4 (i.e., they only guess the target once).
  - a. **Because the listener has the same lexicon as the speaker, they also correctly guess ‘cow’ in this instance with probability 1.**
7. The listener updates their distribution for ‘cow’ as follows:
  - a. Let  $Y$  = the utterance’s vector of probabilities for each gesture
    - i.  $Y = [.5, 0, 0, .5]$
  - b. Let  $Z$  = the listener’s vector of probabilities for object  $m$ 
    - i.  $Z = [.33, .33, 0, .33]$
  - c.  $\text{delta} = \text{learning\_rate} * (Y - Z)$   
 $= .1 * ([.5, 0, 0, .5] - [.33, .33, 0, .33]) = [.017, -.033, 0, .017]$
  - d. Let  $Z'$  = new vector of probabilities for object  $m$
  - e.  $Z' = Z + \text{delta}$   
 $= [.33, .33, 0, .33] + [.017, -.033, 0, .017] = [.347, .297, 0, .347]$
  - f. Renormalize the new probabilities  $Z'$  so they sum to 1:
    - i.  $Z' = Z' / \text{sum}(Z')$   
 $= [.347, .297, 0, .347] / \text{sum}([.347, .297, 0, .347]) = [.35, .30, 0, .35]$
  - g. **The listener’s new distribution for ‘cow’ is [.35, .30, 0, .35].**

These speaking and listening processes can work – i.e., listeners rarely guess wrong – even when there is overlap in objects’ gesture distributions (as in table 1) because internal object distributions with *no* probability over an uttered CC have a squashed kl-div of 1, making their similarity and thus probability of being guessed 0. To illustrate, consider the initial lexicon in table 4a. If a speaker attempted to communicate the 1<sup>st</sup> object and ended up producing an utterance containing the 2<sup>nd</sup> and 5<sup>th</sup> gestures, a listener with the same lexicon and using the process described here would guess the 1<sup>st</sup> object with probability 1 and all other objects with probability 0, *because even the neighboring objects (2 and 7) have no probability associated with one of the uttered gestures*. Communication can only fail when a speaker produces a string in which more than one object has nonzero probability for *all* the uttered gestures. For this lexicon, that could only happen if a speaker produced an utterance with just gestures 1 and/or 2, *or* gestures 5 and/or 6.

Finally, while agents in the Richie et al. (2014) model were paired into interactions randomly (modulo their social network structure), agents in the present simulations were paired up and interacted in the same schedule as in the experiment (see table 1; the basic pattern is: agents A and B interact, then C and D, then AC, BD, AD, BC, repeat from start), with the modification that the schedule of interactions repeated until some preset number of interactions had occurred. Further, whereas agents in the experiment discussed objects in a random order, agents in the model took turns describing and comprehending the objects in numerical order. For example, agent A would describe and agent B would attempt to comprehend object 1, and then they would switch roles. Then, agent A would describe and agent B would attempt to comprehend object 2, and then they would switch roles. This process would continue for the rest

of the objects, after which the next pair of agents in the interaction schedule would converse in the same turn-taking order.

**Table 4a.** An illustration of an initial lexicon of seven objects, each associated with only 6 of 28 possible gestures. Each cell indicates the probability of producing a particular gesture when describing an object. Each object overlaps with two gestures of each of its neighboring objects (see highlighted cells). A kind of semantic space is thus defined, with neighboring objects being semantically similar to the extent that they are associated with similar iconic gestures. Notice that probabilities wrap around the space, as illustrated by object 7.

		Gestures / Conceptual Components																											
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Objects	1	.16	.16	.16	.16	.16	.16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	.16	.16	.16	.16	.16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	.16	.16	.16	.16	.16	.16	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	.16	.16	.16	.16	.16	.16	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.16	.16	.16	.16	.16	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.16	.16	.16	.16	.16	.16	0	0
	7	.16	.16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.16	.16	.16	.16

**Table 4b.** An illustration of how the lexicon in Table 1 evolved in a single simulation lasting approximately 700 pairings. Each cell indicates the probability, written in scientific notation ( $1.09\text{E-}69 = 1.09 * 10^{-69}$ ), of producing a particular gesture when describing an object. Highlighted in orange are the probabilities that are over .1. Three objects are nearly uniquely associated with a single gesture: objects 1, 2 and 5. The remaining objects are still in the process of concentrating probability on a single gesture: owing to the positive feedback in the learning process, a lexicon split over multiple gestures is unstable.

		Gestures / Conceptual Components																											
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
O b j e c t s	1	1.1 4E- 13	2.7 1E- 15	1.0 3E- 06	1.0 0E +0 0	1.4 2E- 09	3.8 9E- 13	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	
	2	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	8.0 5E- 14	8.4 2E- 14	8.7 5E- 16	1.0 0E +0 0	5.8 6E- 16	4.8 1E- 12	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	
	3	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	1.6 4E- 13	7.2 0E- 11	1.7 1E- 01	8.2 9E- 01	4.1 0E- 17	3.6 6E- 17	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	
	4	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	2.3 5E- 01	7.2 6E- 16	5.8 7E- 06	7.6 5E- 01	2.2 1E- 13	7.1 9E- 14	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	
	5	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	1.4 6E- 17	4.5 1E- 12	9.9 8E- 01	2.1 9E- 03	1.8 5E- 15	6.5 3E- 09	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	
	6	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	2.9 6E- 16	4.3 1E- 13	4.4 1E- 01	5.5 9E- 01	1.6 4E- 15	1.8 2E- 12	0.0 0E +0 0	0.0 0E +0 0	
	7	7.8 9E- 16	2.2 1E- 15	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	0.0 0E +0 0	4.5 5E- 13	4.1 5E- 07	3.4 1E- 01	6.5 9E- 01	

## Model behavior

The parameter space for the model is sizable: there are parameters for the number of agents, objects, conceptual components/gestures, the overlap between neighboring objects, and the learning rate. As yet, we have only explored a small portion of this space. All simulations reported here use 4 agents, discussing 7 objects each initially associated with 6 gestures, two of which are unique to an object (as in table 4a), with a learning rate of .1. Future work with this model would of course need to explore the parameter space further, to understand, for example, how much the present findings of simultaneous conventionalization and reduction depend on the present parameter settings.

Figure 9 illustrates the evolution of jaccard index (the measure of conventionalization), communication success rate, and utterance length in a single simulation of the new model, and Figure 10 shows the same quantities averaged across 20 simulations. The figures illustrate two of the *Key Aspects* that the model captures: agents simultaneously conventionalize, reduce, and increase success at the task (desideratum 5), but despite low initial conventionalization, communication success starts quite high (desideratum 6).

Why do conventionalization, reduction, and increasing success (co-)occur in the way that they do? Conventionalization is initially low because probability is dispersed across several gestures for a given object, and agents will usually not randomly select the same gestures for the same object. At the same time, communication success starts relatively high because (1) speaker and listener share concentration of probability on a set of gestures unique to the target object, making it likely that the speaker's utterance is closer to the target object in the listener's lexicon than any other object, and (2) the speaker's sample-until-recognition process implies that each generated utterance comes fairly close to approximating the (distinctive) distribution that generates it. Finally, utterance length is initially high in this model because speakers often need

to utter multiple gestures before they recognize/are confident in their utterance – their first gesture often overlaps with other objects, so the speaker doesn't recognize the target object and must keep uttering gestures. For example, imagine a pair of agents interacts with the lexicon in table 3a, and the speaker attempts to communicate object 1. The speaker randomly produces gesture 2, doesn't randomly guess object 1 (because they are equally likely to guess object 5), so then randomly utters gesture 1, but now happens to guess object 1 (even though object 5 is still possible). The listener then likewise attempts to guess the target, but they incorrectly guess object 5. Later, when the listener attempts to communicate object 1, they might produce many different combinations of gestures aside from {2,1}; the probability of these two agents (let alone all four agents) producing the exact same utterance for this object is low.

Then, as agents interact, and as listeners move towards their speakers, the particular utterances speakers (probabilistically) produce are reinforced, making them even more likely to be produced in the future (a positive feedback effect). Because agents move towards one another, these reinforce utterances will be shared across agents, constituting conventions; further, different runs of the model produce different conventions, owing to the random sampling, mirroring the quad-specific conventions in the experiment (desideratum 4). The listener's update procedure transfers probability from non-uttered gestures to uttered gestures, making the probability distributions less diffuse (lower entropy, in the information-theoretic sense). Because the probability distributions are less diffuse, and agents are unlikely to reinforce the same gestures for different objects (because the choice of which gestures to reinforce is random, and because the initial gestural overlap between objects is limited), different objects' distributions overlap less, making comprehension – by the speaker or the listener – easier.

To put this last point a little bit differently, evolved lexicons, in which objects' distributions overlap less, mean speakers are less likely to produce ambiguous gestures. Ambiguous gestures are unlikely because putting probability on multiple objects for the same gesture is unstable. Suppose a gesture is associated with two objects with equal probability. When a listener is trying to infer the target object, the choice of object will be 50-50, but they will happen to choose one, and now the association between that gesture and that object will be reinforced. Now, the next time this gesture is produced, the listener will be slightly biased towards the object that they previously inferred. Hence, through positive feedback, the gesture's association for one object will increase, while the association with the other object will decrease (because other gestures will be rewarded for that object). Finally, because the speaker comprehends their own utterances more easily, they tend to produce shorter utterances.

Table 4b illustrates a lexicon having completed the process just described. (It is a single agent's lexicon, but the other agents' lexicons will be nearly identical, owing to the conventionalization process, so showing just this agent's lexicon will suffice.) The main feature is that each object is now associated with one object, with probability rounding up to 1, and every other gesture with probability 0 or approaching 0.<sup>11</sup> Further, different objects conventionalized different gestures. Now, when this agent wants to convey object 1, they are virtually guaranteed to select gesture 4 and understand their own utterance immediately, without having to utter additional gestures. Since their listener shares this lexicon, they are also virtually guaranteed to correctly guess object 1, and when it is their turn to convey object 1, they will also select gesture 4. Thus, short utterances and nearly perfect communication and conventionalization obtain at this point.

---

<sup>11</sup> The fact that each row in the evolved table does not add up to 1 can be attributed to limits in how precisely floating point numbers can be represented.

Finally, there is the question of whether this model reduces more often after failed communication events compared to successful events, as happened in the experimental data (desideratum 7). Subjecting simulation results to analyses similar to those in the experimental section, we do find evidence that the model reduces more often after failure. We simulated 20 runs of the model using the same parameters as in figure 9, and for each run, counted the number of events of reduction and the number of events of nonreduction relative to the agent's last utterance of that object. For each of those events, we also checked whether the agent's last utterance of that object was a success or failure. We then computed each simulation's odds of reducing following success, and odds of reducing following failure. We found that 18 of 20 simulations showed greater odds of reducing following events of failure than events of success ( $p = .0004$  by a binomial test); across simulations, on average 4.88% of events of success were followed by reducing, compared to 3.73% of events of failure. This relationship might be understood as follows: in the limit, communication success is (near) perfect, and utterances are one gesture long and cannot be reduced further. Thus, as successful events continue accruing without attendant reduction, the rate of reduction following successful events is low. This, in turn, guarantees that reduction is more likely after failure.

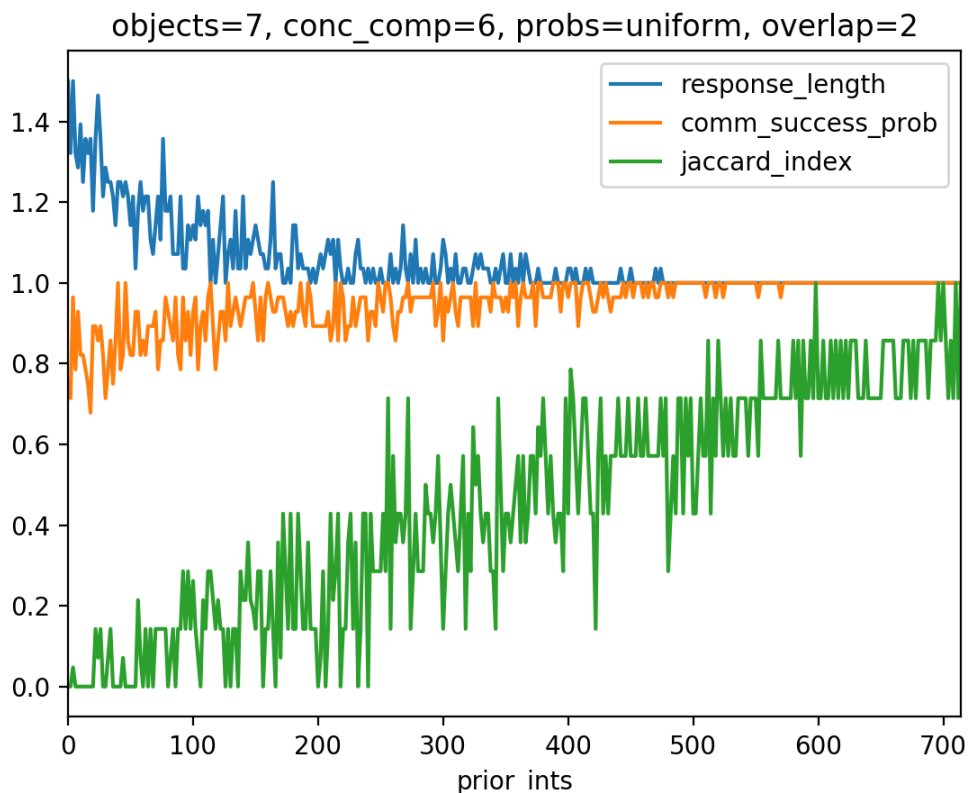
A computer-aided function exploration suggests that this correlation between reduction and failure reflects a necessary, mathematical truth about the relation between two monotonic functions  $f(x)$ , in *any* span of  $x$  (which here represents time). That is, reduction seems guaranteed to occur at a higher rate after failure than after success as long as the rate of reduction monotonically decreases and the rate of communication success monotonically increases over time (as we tentatively claim they do in the experiment and in the model, see figures 11 and 12). We tested this with reduction and communication trajectories of different functions as follows.

First, assume that the probabilities of reduction and communication failure both follow functions like  $e^{-x}$ , the exponential decay function. At each time point  $x$ , we multiply the probability of reduction times the probability of communication success (or failure) to obtain the joint probability of an event of reduction, following an event of success (or failure). We sum these probabilities across all time points and then divide by the sum of the probabilities of success (or failure) at each time point, to yield the rates of reduction following success (or failure). We found that reduction occurred at greater rates following failure under several different functions of reduction and failure:

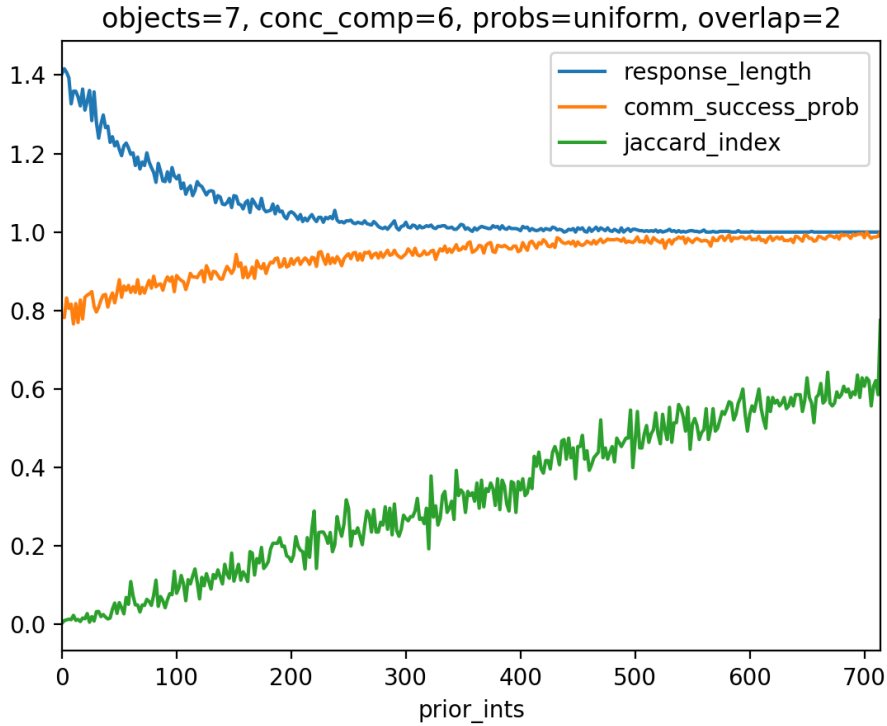
- Rate of reduction and rate of failure both following  $e^{-x}$
- Reduction following  $2^{-x}$  and failure following  $3^{-x}$  (i.e., communication failure decreasing faster than reduction)
- Reduction following  $3^{-x}$  and failure following  $2^{-x}$  (i.e., reduction decreasing faster)
- Reduction following  $.5 * 3^{-x}$  and failure following  $.5 * 2^{-x}$  (i.e., different exponential decay curves, with reduction decreasing faster than failure).

To repeat, this analysis suggests that the reduction~failure correlation that we found follows from general mathematical properties and may not be particularly specific to our model or data. From this standpoint, correlation of reduction with failure is sensible, but it is also worth thinking back to how agents in the model behave under success and failure, and how this compares to listener-oriented theories of speaker and listener behavior. In listener-oriented theories, a speaker adjusts their productions based on the perceived success or failure of communication with the listener. In our model, however, agents do not reason about each other's understanding at all – they, and only *the listener* at that, only modify their future behavior to better align their own utterances for each object with what they guess their interlocutor is talking

about. In that sense, our model embodies a theory rather different from listener-oriented theories of why language users reduce (the count of morphemes in) their utterances.



**Figure 9.** Evolution of a single representative simulation of the model, where agents discuss 7 objects, and start with probability distributed across 6 gestures (conceptual components) for each object. Each object shares 2 gestures with its neighboring object, meaning that only 2 gestures are unique to a particular object; similarly, in the experiment, about 40% of the gestures ever used for an object were unique to that object. The learning rate gamma was set to .1. On the x-axis is number of network-wide interactions (the same measure of time used in the experiment), and the y-axis is shared between jaccard index (conventionalization), communication success, and response length. All three quantities co-evolve in a way similar to that of the experiment – conventionalization rises from near floor to (near) ceiling, communication success starts high but improves even further, also approaching ceiling. Utterances are initially more than 1 gesture long, on average, but shorten over time.



**Figure 10.** Averaged behavior of 20 simulations of the model, where agents discuss 7 objects, and start with probability distributed across 6 gestures (conceptual components) for each object. Each object shares 2 gestures with its neighboring object, meaning that only 2 gestures are unique to a particular object. The learning rate gamma was set to .1. On the x-axis is number of network-wide prior interactions (the same measure of time used in the experiment), and the y-axis is shared between jaccard index (conventionalization), communication success, and response length.

## Conclusions of modeling

To summarize the results of this section, our model demonstrates that conventionalization, reduction, and communication success – and particularly the relationships among them -- can be explained by positing iconicity in the gestural modality, and simple probabilistic mechanisms of gesture/language production, comprehension, and learning. Further, the correlation between conventionalization and reduction observed in the lab and in naturalistic settings requires no links of speaker-oriented or listener-oriented varieties. Rather, both conventionalization and reduction can result from agents (human or simulated) limiting their associations between gestures and objects to a narrower subset of gestures for a given object, via positive feedback of the gestures that are probabilistically produced for an object at early stages

of the group's interactions. This narrowing makes agents more certain of what to say when attempting to communicate an object, which translates into more concise expressions, and makes the listener more certain of the target object's identity.

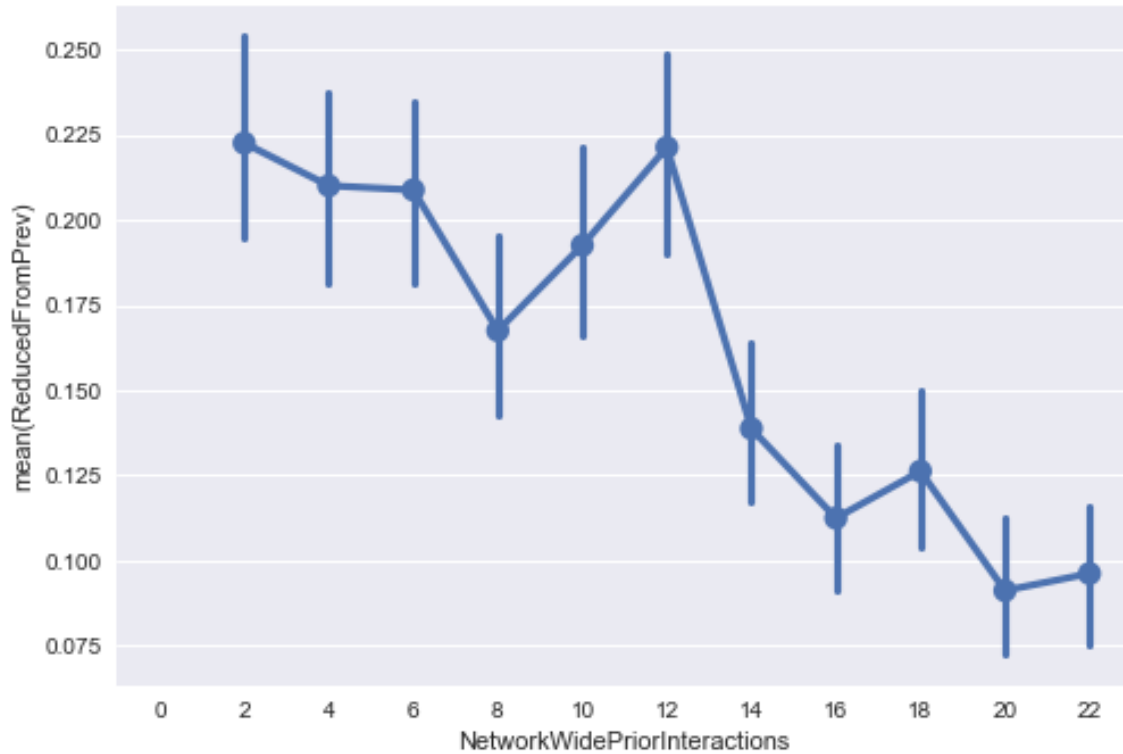


Figure 11. The rate of reduction in the experiment. The mathematical analysis described at the end of section *Model Behavior* was predicated on a monotonically decreasing rate of reduction. It is unclear whether the rate of reduction in the experiment meets this assumption – the rate is monotonic but for sharp increases at ~10 and 12 network-wide prior interactions. Given the large, largely overlapping error bars between 8, 10, and 12 interactions, however, this apparent departure from monotonicity may simply be noise.

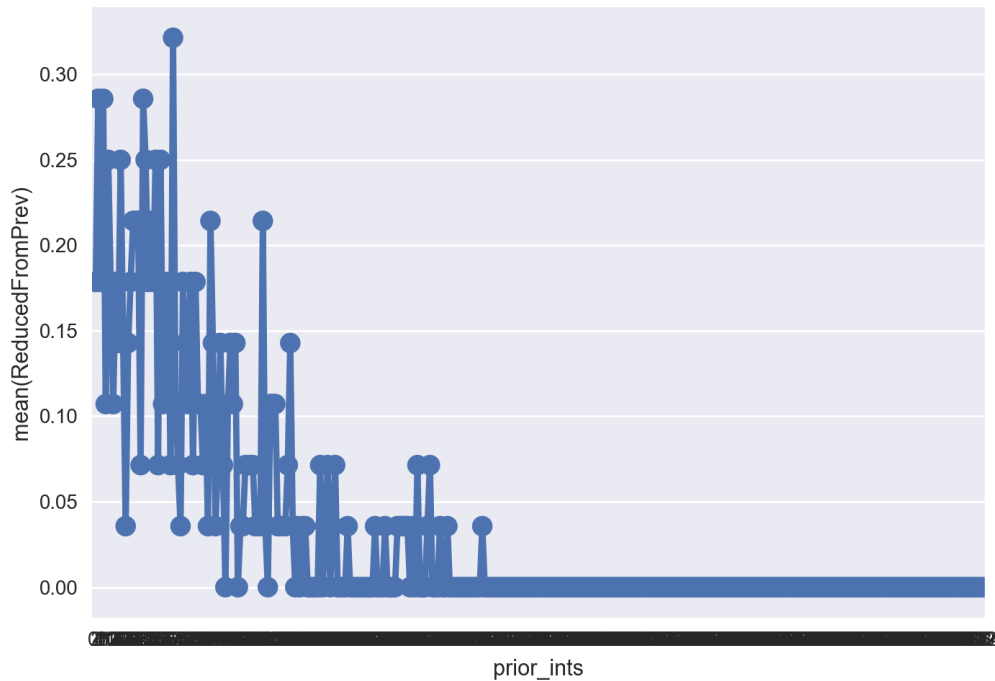


Figure 12. The rate of reduction in a single run of the model. The mathematical analysis described at the end of section *Model Behavior* was predicated on a monotonically decreasing rate of reduction. Departures from monotonicity in this figure are due to noise.

## Chapter 5 - Discussion

How do people invent communication systems? In naturally emerging languages, this process seems to involve both conventionalization – people creating *shared* systems of form-meaning mappings – *and* reduction – linguistic forms being shortened over time (Osugi et al., 1999; Meir et al., 2010; Richie et al., 2014). How do conventionalization and reduction (co-)occur in language emergence and change? Previous theories of conventionalization, often embodied in computational models, have usually gone untested against empirical data. Further, while iconic forms have been observed to have an important role in language emergence (Meir et al., 2010; Richie et al., 2014), this role has not been formalized in any computational models of

language emergence. Prior theories of reduction, on the other hand, have usually not been concerned with reduction in emerging languages, let alone signed languages, but rather in mature language change and short-term speaker choices. Finally, theories of the relationship *between* conventionalization and reduction have been vague and mostly concerned with reduction of phonetic substance rather than count of morphemes (or other discrete elements).

This dissertation attempted to fill in these gaps, with an experiment requiring participants to negotiate a gestural communication system, and a computational model of conventionalization and reduction in that experimental setting. We found that, as in natural settings, participants in our experiment conventionalized and reduced communication systems simultaneously. Crucially, these systems were self-organized, and the ultimate systems possessed a degree of arbitrariness: different groups conventionalized different systems, each of which accomplished the task of communication comparably. Further, we found that participants did *not* reduce at higher rates after communication success, but rather after failure, a finding inconsistent with certain listener-oriented theories of reduction (and the link between conventionalization and reduction). Then, in a computational model, we developed an alternative theory of conventionalization and reduction. In this model, agents (human or simulated) start with associations between objects and subsets of gestures iconically related to a given object. Then, *via probabilistic gesture production and positive feedback for those gestures that happen to be uttered for an object*, agents narrow their associations to smaller subsets of gestures for a particular object. This narrowing makes agents within a group more likely to produce the same gestures for a given object – this constitutes conventionalization. The narrowing of gesture subsets for a particular object also makes agents more certain of what to say when attempting to communicate an object. This greater certainty translates into more concise – that is, reduced – expressions, and also makes the listener more

certain of the target object's identity, achieving greater communication success. In sum, the novelty of this theory is that conventionalization and reduction can be captured by probabilistic form-meaning mappings, language production dependent on a notion of informativeness (understanding one's own utterance), and a single learning mechanism in listeners, whereby listeners align their probabilistic associations with speaker's utterances.

In the remainder of the discussion, we (1) relate these findings to existing theories of language emergence, conventionalization, and reduction, (2) discuss limitations of the present work, (3) describe future directions, and (4) summarize the findings and their implications.

### **Relation to existing theories of language emergence, conventionalization and reduction**

We first note how our model is consistent with past modeling frameworks of language emergence, and then discuss our model's novelty.

As noted by Spike et al. (2016), the emergence of (human) communication systems has been observed in myriad experimental and naturalistic settings, and modeled in equally numerous ways, yet there is little consensus on the mechanisms driving the emergence of these communication systems. By far the strongest theoretical consensus seems to be that conventionalization of communication systems can be self-organized among agents lacking common knowledge (all agents are aware of a set of propositions, each agent knows that every other agent also knows those propositions, and so on recursively ad infinitum). Our model, like prior ones, also demonstrates this. Similarly, our model possesses the three qualities of successful models of language emergence that Spike et al. found in several families of such models. In the interest of continuing the theoretical integration begun by Spike et al., we now attempt to understand our model in terms of their three requirements for the emergence of conventional signaling: (1) creation and **transmission of reference**/referential information, (2)

**anti-ambiguity bias**, and (3) **information loss**.

Spike et al.'s (2016) first requirement is that speakers must have some way of propagating information about the intended referent to the listener. For example, in some models, speakers point at their intended referent (sometimes after failed communication) while in others the possible referents in a communicative act are only a subset of the wider universe of referents. Our experiment and model use neither of these approaches: listeners get no disambiguating information about intended referents, and the set of possible referents is the same on every trial. How do our participants and agents propagate referential information, then? They do so through iconicity: because participants and agents start with (nearly) equal iconic mappings between gestures and objects, the use of a particular gesture or set of gestures *already provides a great deal of referential information to the listener*. Thus, our model still satisfies Spike et al.'s first requirement, albeit in a possibly novel, unexpected way. In the next section we discuss this feature and its implications for generalizability of our work in greater depth.

Our model satisfies the second and third requirements in a more straightforward fashion. The second requirement is that there be some bias against ambiguity or homonymy – agents can not conventionalize the same form for every possible meaning, or else communication will not reliably succeed for different meanings. As discussed in the modeling section, the agents do typically conventionalize different forms for different meanings, because (a) agents are unlikely to randomly select the same gestures to reinforce into conventions for multiple objects, and (b) ambiguity is evolutionarily unstable: the positive feedback in the speaking-listening-learning loop ensures that, when one gesture has equal probability with two or more objects, probability will eventually concentrate on one object at the expense of others. To put this differently, the anti-ambiguity bias stems from the interaction of independently necessary agent components

(speaking, listening, learning), and requires no additional agent cognition. Unlike models where agents have an *explicit* anti-ambiguity mechanism (i.e., periodically removing homonyms from the lexicon, or lateral inhibition among homonyms, e.g., Smith, 2002; Steels & Loetzsch, 2012), our anti-ambiguity bias is a side-effect of the independently necessary components of comprehension and learning.

The third requirement is that there be some form of information loss, through, e.g., agents simply forgetting past form-meaning associations, or older agents being replaced by younger ones. Spike et al.'s (2016) intuitive explanation for this is that information loss reduces the impact of the early, disordered state of the communication system on the development of an optimal system. Our model essentially takes the forgetful agent approach: agents have no memory of their past interactions or lexicons that delivered them to their current lexicon; they remember only their current lexicon itself, which encodes minimal information about the agent's history (because there are an infinite number of possible evolutions to any given lexicon).

The above discussion illustrates how our model is consistent with past theoretical development of language emergence. However, our model is novel in a number of respects. As mentioned in the introduction, Spike et al. (2016) point out that different empirical settings of communication could possibly engage different cognitive mechanisms for inventing communication systems. Despite this, as also mentioned in the introduction, integration of modeling and empirical investigation in language emergence is relatively rare. Our work here, however, is such an attempt at integration: our empirical setting constrained our model in a number of ways, including (1) the role of iconicity, (2) the lack of reliable, explicit communicative feedback, and (3) absence of evidence of a listener-oriented link between conventionalization and reduction. Future work could examine how changes to our setting

required – or not – appeal to different cognitive mechanisms for language creation. Our model is also novel in its accounting for conventionalization *and* reduction in language emergence. We now turn to relating our work to theories of reduction, and the relationship between conventionalization and reduction.

First, our experimental finding that failure correlates with reduction is surprising given past findings that language users enhance their utterances following (perceived) failure to communicate (Buz, Tanenhaus, & Jaeger, 2016; Schertz, 2013; Stent, Huffman, & Brennan, 2008). Two are at least two different possible reasons for this unexpected finding in our study. First, consecutive events of talking about a given object are minutes apart, and utterances about many other objects intervene. This may make it difficult for people to remember their perception of the success/failure of the previous attempt to communicate about a given object. Second, the signals of success/failure given by comprehenders might not be strong enough. Our experimental method prevented people from simply saying (in English) “I don’t understand”, nor could participants see their interlocutor’s choice of target, nor did experimenters give corrective feedback. Participants could only rely on gestures and facial expressions to signal their comprehension (see the following section on limitations of the present work for further discussion of this design choice). In other experiments where participants enhance following failure, however, incorrect target choices are explicitly signaled to the speaker (e.g., Buz et al., 2016), and enhancement followed immediately after perceived errors as speakers attempted to repair communication in the same instance of communication (Stent et al., 2008). These design differences relative to previous work may account for our unexpected finding that reduction correlates with communication failure.

The experimental and modeling findings that failure correlates with reduction isn’t

necessarily inconsistent with the listener-oriented theory – the correlation is perhaps just a necessary mathematical truth – but the model we develop here does still embody a theory of conventionalization and reduction that is somewhat different from prior listener-oriented, and possibly speaker-oriented, theories of reduction and the link between conventionalization and reduction. That is, our model is different from listener-oriented theories of language production and reduction in that agents do not reason about or even receive information about each other’s comprehension. Put differently, agents are not modulating their expressions based on the perceived comprehension of their interlocutors – listeners only adjust their future expressions based on their own comprehension of a speaker’s utterance.

Whether our model is different from speaker-oriented theories is a bit less clear. Speaker-oriented theories of reduction broadly hold that reduction results from language production becoming ‘easier’ (Jaeger & Buz, 2016). In our model, agents’ gradual concentration of probability on smaller subsets of gestures leads to speakers’ faster satisfaction with their own utterance. One could possibly construe this as agents having ‘an easier time’ producing language. This possibility of construing our model as a speaker-oriented/production ease account of reduction and its link to conventionalization is particularly intriguing given recent skepticism by reduction theorists (e.g., Jaeger & Buz, 2016) that production ease accounts could explain morphological reduction (other cases including, e.g., omission of the optional complementizer ‘that’ as in “She knows (that) he’ll cook dinner”).

If the model can be construed as a speaker-oriented / production ease account, it is worth considering how this account compares to accounts of the link between conventionalization and *phonetic* reduction. As noted in the introduction, conventionalization is often accompanied by reduction of phonetic variables like vowel duration and signing articulatory space. To reiterate,

theorists typically explain this link as follows: conventionalization causes forms to be used repeatedly, which causes them to become automated or routinized, which causes them to be shortened (Bybee, 1999, 2006; Namboodiripad et al., 2016).<sup>12</sup> This theory is actually slightly but importantly different from our model: in our model, conventionalization and reduction are caused simultaneously by a single learning mechanism, whereas in cases of phonetic reduction, conventionalization (purportedly) *causes* reduction. Thus, both phonetic reduction and the morphological reduction in our study could conceivably both be considered ‘speaker-oriented’, but their causal networks are somewhat different. Still, the possibility of a somewhat unified account of reduction phenomenon in emerging language conventionalization is interesting.

Finally, as mentioned in the introduction, part of the import of reduction phenomena is that they are often understood as making language systems ‘efficient’, by saying no more than necessary for communication success. However, as noted in other critiques of functionalist explanations of language structure and other behavior (e.g., Haspelmath, 1999; Scott-Phillips, Dickins, & West, 2011; Richie, 2016), to say that a language has a certain feature (here, short utterances for predictable information) *because* it is functional, requires showing that the language evolved that feature *under pressures for that function*. Our model, however, shows that a lexicon can evolve from longer expressions to shorter, more efficient expressions, even when there is no explicit optimization of any notion of ‘effort’. Indeed, agents in our model never even explicitly represent utterance length or any other effort-related quantities. The agents’ creation of efficient lexicons ‘falls out’ or emerges merely from speakers probabilistically producing

---

<sup>12</sup> In the phonetic cases of reduction that Bybee (1999, 2006) and Namboodiripad et al. (2016) were concerned with (‘going to’ → ‘gonna’), automation/routinization purportedly leads to greater coarticulation, overlap and ultimately elimination of segment boundaries and hence reduction in count and duration of segments. This could in principle lead to morpheme elimination of the kind that we were interested in here, but it would manifest as gradual rather than sudden disappearance of gestures. We have not conducted a phonetic-level coding of the data that would allow us to detect this, but we suspect gestures are eliminated without phonetic reduction precursors.

utterances until they recognize their target object, and then listeners updating their lexical entry towards that utterance.

### **Limitations of the present work**

Several considerations and caveats are necessary when generalizing the present findings. First, as with all experimental semiotics studies, the experimental setting is minute in scale and complexity compared to naturalistic language emergence – our participants only needed to communicate about 25 basic objects and didn't have to describe *events*, and only interacted over a handful of hours, whereas natural language emergence unfolds over years and decades (Senghas et al., 2004; Sandler et al., 2011). That said, we believe that our participants were engaging similar cognitive mechanisms as those humans would use in natural language emergence settings.

Another aspect of our experimental procedure perhaps lacking in ecological validity is the fact that participants did not receive explicit, reliable corrective feedback from experimenters *and* could not point to the referents of their utterances, as people do in naturalistic settings. That the experiment lacked experimenter feedback is perhaps not so bad – in the real world there is no omniscient third party guiding communication. But the impossibility of speakers using disambiguating deictics (before or after the listener attempted comprehension) does depart somewhat from natural settings. While sometimes people *can't* point at the things they are talking about (for example, when their hands and body are occupied with other things), when they can point, they often *do*. A chef in a kitchen might point to an unusual ingredient while asking a novice busboy to get it. Similarly, a parent might point to an object while naming it for their child (and such 'high-quality' learning events are thought to be an important driver of child word learning; Cartmill, Armstrong, Gleitman, Goldin-Meadow, Medina, & Trueswell, 2013). The omission of *both* reliable corrective feedback *and* disambiguating deictics means that

participants have no way besides iconicity of understanding each other. In some ways, this makes our participants success even more striking, but as discussed, it makes our experiment less applicable to certain real-world settings.

There is a related concern regarding our analysis of the relationship between communication success and reduction. To repeat, the listener-oriented theory we tested holds that speakers reduce (or conversely, enhance) their utterances when their communication partner(s) indicate ease (or conversely, difficulty) with comprehension. However, we did not directly measure listeners' ease/difficulty of comprehension, but rather made the assumption that successful or failed communication events were accompanied by listeners indicating (intentionally or not) their ease or difficulty to their interlocutors, and then related communication success/failure to future reduction. This assumption strikes us as reasonable, but clearly a more stringent approach would be to (a) code video data for listener signals of understanding/confusion, and (b) correlate such signals with future reduction/enhancement. This is one possible route for future work.

Relatedly, the present experimental analyses are far from an exhaustive test of listener-oriented theories. For example, such accounts of reduction often hold that linguistic forms that carry little information are those that are likely to be reduced or entirely eliminated (Piantadosi et al., 2011; Mahowald et al., 2012; Jaeger and Buz, 2016). Frequent forms are often those that are predictable, and thus can be eliminated. This possibility could surface in our experimental data as follows: the gestures that a participant eliminates for a given item at time  $n$  would be more frequent across items in previous productions, compared to the gestures that the participant produces at time  $n$ . One difficulty in implementing this analysis, however, lies in deciding *what* window of past productions to include for calculating gestures' frequencies. Should the analysis

include all available productions, i.e., since the beginning of the experiment, or only those productions since the last time the participant produced the item in question? And should only the producer in question's productions be considered, or should interlocutors' as well? Interlocutors' productions could conceivably contribute to a speaker's estimates of form frequency. These are questions we leave to future work.

Finally, there is the question of the extent to which our experiment and modeling can only generalize to systems in the same modality that we studied: gesture and sign. Gesture and sign afford considerable iconicity, and our experimental setup and computational model critically relied on this iconicity, as neither allowed for any other channel of transmission of referential information from speaker to listener (in the form of, e.g., disambiguating deictics or restricted sets of possible referents). As pointed out by Spike et al. (2016) and discussed in the previous section, transmission of referential information is critical for the emergence of conventional signaling systems, and iconicity served this purpose in our experiment and model. One might think that spoken languages do not contain comparable levels of iconicity that allowed a model like ours to work.

There are a few responses to this concern. First, sign languages are natural languages as expressively powerful and structurally complex as spoken languages (Sandler and Lillo-Martin, 2006), and used by tens of millions of people around the world (World Federation of the Deaf, 2016), and are therefore objects worthy of study in their own right. That is, even if the present conclusions generalized only to sign languages, we believe that would still be a significant theoretical development. Second, spoken languages contain a degree of iconicity, too (for reviews, see Perniss, Thompson, & Vigliocco, 2010; Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015). In addition to onomatopoeia (e.g. 'bang' for a gunshot, or 'splash' for an

object hitting water), there is the well-known ‘kiki’-‘bouba’ effect (people tend to think a sharp object is named ‘kiki’ and a round object ‘bouba’). Further, the fact that iconic forms in sign language are often rendered arbitrary by historical change (Frishberg, 1975) might suggest that iconicity in emerging spoken languages has been obscured somewhat by language change. Though of course this is a theory that may be difficult to test with actual, emerging spoken languages, experiments suggest that emerging communication systems in the vocal channel can utilize iconicity (Perlman, Dale, & Lupyan, 2015). Thus, to the extent that (young) spoken languages are iconic like signed languages, there should be some generalizability from our findings to spoken languages.

However, there are other caveats concerning the role and treatment of iconicity in the present work. For one, the model obviously oversimplifies iconicity: it is not the case that all objects one might want to talk about have a fixed amount of semantic overlap and non-overlap with other objects. Further, different people clearly have different semantic knowledge, and this can manifest in how their gestures iconically represent concepts. For example, in Japanese Sign Language ‘rock’ is signed with a throwing motion (purportedly because Japanese fisherman weight down their nets with rocks), while in American Sign Language ‘rock’ is signed by tapping the dominant hand onto the other hand (possibly representing the hardness of rocks). Similarly, children and adults differ in the extent to which they recognize and leverage iconicity in gestures, as perceiving iconicity is somewhat demanding of perceptual-conceptual mapping resources (Magid & Pyers, 2017). Of course, our experiment used adults who have an easier time encoding and decoding iconicity than children. To the extent that iconicity is necessary for bootstrapping a communication system, children’s limited ability to use iconicity may mean that they have less of a role in conventionalizing lexical items/referring expressions, whereas prior

work suggests they have a special role in shaping certain aspects of grammar (see Senghas & Coppola, 2001; Senghas, Kitz, & Ozyurek, 2004). Future work could (1) enrich the model's iconic/semantic representations based on, for example, human-generated feature norms, and (2) investigate how inter-agent differences in initial semantic representations, and use of iconicity, influence language evolution. As discussed earlier, our models' agents sharing an initial semantic space was critical for successful evolution of the system, so some amount of inter-agent consistency will be necessary for any language emergence operating as it does in our model, but the precise amount is currently unknown.

### **Future directions**

In the introduction, we argued that conventionalization (and reduction) may be key to understanding the emergence of phonology and grammaticalization. We now reconsider these phenomena in relation to the present work.

Regarding phonological combinatoriality, we must reiterate that the current experimental analyses and modeling did not investigate gestures at the phonetic level. That is, we did not code gestures for qualities like handshape, location, or movement (parameters key to sign language phonology; Sandler & Lillo-Martin, 2006). We only represented and modeled gestures in terms of their iconic meanings. In that sense, the present work by itself does not touch on the emergence of phonology. However, our participants very well might have invented phonological systems. That is, participants may have been conventionalizing not just particular gestures for particular objects, but also a limited inventory of handshapes, locations, movements and so forth with which to compose those gestures. How could such phonological categories be detected in systems such as the ones our participants created? In mature, stabilized languages, phonological categories are often determined by finding minimal pairs that differ only on particular segment or feature. For example, the existence of the words 'pin' and 'bin' in English is usually taken to

mean that /p/ and /b/ constitute separate phonological categories (phonemes). The same method is used in sign languages: the signs for ‘mother’ and ‘father’ are a minimal pair differing only on location (‘father’ is articulated at the forehead, and ‘mother’ at the chin). In principle, then, this same approach could be applied to gestural systems such as ours. However, the small number of meanings to be conveyed (25) relative to the number of possible articulatory configurations (i.e., combinations of handshapes, locations, movements, etc.) to convey these meanings makes it unlikely that two different gestures are adjacent to one another in this articulatory space (i.e., are minimal pairs). For this reason, the minimal pair method of determining phonological categories may be insufficient for the present experiment.

Alternatively, phonological categories could possibly be assessed using motion-capture to yield fine-grained information about gestures in space. If participants are creating a limited inventory over the course of the experiment, then we should observe a gradual clustering of forms in gestural space. That is, initially, participants’ handshapes, locations, and movements should be relatively more uniformly distributed throughout the possible configurations of handshape, location, and movement. Over time, as participants selected a limited inventory of these, we should see multi-modal distributions, with modes corresponding to the prototype for each category of the limited inventory of articulatory features.

However, even if a method of detecting emergence of phonological categories could be devised, the present experiment might not have sufficiently encouraged phonological development. For one, experimental evidence suggests that allowing iconic forms, as the present experiment did, inhibits the emergence of phonology (Roberts, Lewandowski, & Galantucci, 2015). Second, the emergence of phonology is often argued to be a solution to the problem of lexical contrast posed by large lexicons (Hockett, 1960; Pinker & Jackendoff, 2005; Nowak &

Krakauer, 1999; Zuidema & de Boer, 2009; Wedel, 2012).<sup>13</sup> Thus, to the extent that (a) the emergence of phonology is *driven* by large lexicons, and (b) the present experiment did not involve communicating a large number of meanings (25, cf. the tens of thousands of meanings typically encoded by lexicons of mature natural languages, Pinker & Jackendoff, 2005), participants in our experiment may not have been encouraged to develop phonology.

Regardless of the degree of phonology in the present experimental data, simultaneously modeling conventionalization, reduction, *and* the emergence of phonology would seem to require a model of substantial complexity. Such a model would need to simultaneously represent iconicity, as we have done here, but also the phonetic qualities along which iconic gestures vary. It is unclear the precise form such a model would take, but it might be obtained from a synthesis of the present model with prior agent-based models of phoneme inventory evolution, which also involve populations of agents speaking, listening, and learning from one another (e.g., de Boer, 2000; Wedel, 2012).

The import of our work for grammaticalization is similar. Our participants only needed to convey objects and not events, so the gestures they produced tended to represent concrete meanings rather than the abstract meanings typically encoded by grammatical elements. That is, participants had no need to encode concepts like person (e.g., I vs you), tense (e.g., past vs present), aspect (e.g., ‘I run’ vs ‘I am running’), mood (e.g., possibility vs requirement), and relations which are often encoded (at least in English) with prepositions (‘in’), complementizers (‘because’), and relativizers (‘who’ as in ‘the man who sleeps’). Indeed participants only very rarely represented such concepts – anecdotally, the only potentially grammatical forms we

---

<sup>13</sup> This argument takes various forms. One argument is that the problem of distinguishing large numbers of holistic word forms is avoided by concatenating such forms (i.e., ‘cat’ is a concatenation of /c/, /a/, and /t/. A different argument is that combinatorial phonology allows words to be ordered in a list like a phone book, which allows quicker lexical retrieval. Many such arguments are not mutually exclusive.

observed included gestures representing number (to indicate the number of objects on a card) or negation (as in, NOT-LARGE-ROUND, see table 2a) and points to oneself or to one's interlocutor to indicate gender. To this extent, our experiment is unlikely to have produced grammaticalization phenomenon, and our model certainly was not designed to capture such.

This is not to say that the present paradigm and model cannot be useful for contributing to understanding of grammaticalization. Our paradigm could be modified so that participants talk not about objects, but events, as in similar experimental semiotics work in the gestural modality (e.g., Carrigan, 2016). Comparing our data to such a new dataset, different from ours only in the things discussed (objects, vs events involving those objects), might be useful in gauging how describing events drives grammaticalization more than does merely describing objects. Modeling how conventionalization (and reduction) drive grammaticalization may present more of a challenge. For one, in contrast to the emergence of phonology, there are very few models of grammaticalization from which to draw inspiration. One of the few was devised by Tabor (1994). By training an artificial neural network on sentences generated by a probabilistic context-free grammar, and then changing the probabilities of certain rules in the grammar and continuing to train the network, Tabor was able to model the reanalysis (a route of grammaticalization) of the phrase 'be going to' from an indicator of motion (as in "I am going to the store") to a indicator of futurity (as in "I am going to buy milk"). Other models of grammaticalization are not computational models of psycholinguistic processes, but statistical models of lexical change in historical corpora (e.g., Sagi, Kaufmann, & Clark, 2011 on 'do' undergoing semantic broadening, a process often accompanying grammaticalization). Importantly, none of these is a model of interacting humans, attempting to communicate with one another. Thus, the way to integrate prior models of grammaticalization with the current dissertation's model is not entirely

clear. One possibility, only vaguely specified at the moment, is that the agents in the present model could be replaced with artificial neural networks which not only conventionalize with other agents, but also reanalyze – that is, grammaticalize – their linguistic input much like the networks in Tabor (1994). One challenge for such models is to explain how, cross-linguistically, events of language use and learning grammaticalize forms for particular meanings, like motion or possession (as in English ‘I have goats’), into markers for particular grammatical functions, like futurity or perfective aspect (‘I have owned goats’; Dahl & Velupillai, 2013).

## **Conclusion**

The present dissertation investigated the emergence of communication systems – particularly, systems that conventionalize and reduce – among interacting adults and simulated agents. The experimental and modeling work both suggest a new understanding of the relation between conventionalization and reduction. Whereas existing theories of the link between conventionalization and reduction rely on either (1) vague notions of automaticity or (2) agents explicitly optimizing for communicative efficiency, our work suggests a different account: interacting language users evolve an iconic world-referent correspondence system into a system that is conventional and reduced (i.e., efficient) in a social context by adhering to informative utterances and emulating one another. While this view of communication systems as emergent results of self-organizing populations is common in language emergence theories, concrete specifications inspired by detailed empirical observations have been uncommon. We hope future work exploring the self-organization of language emergence will continue to integrate experiments, naturalistic observation, and computational modeling, as we have done here. Such an ecumenical approach will be fruitful in further exploring deeper issues in language

emergence, including phonological combinatoriality and grammaticalization.

## References

- Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*, 19(3), 505–512.  
<http://doi.org/10.3758/s13423-012-0233-y>
- Baronchelli, A., Gong, T., Puglisi, A., & Loreto, V. (2010). Modeling the emergence of universality in color naming patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 107(6), 2403–7. <http://doi.org/10.1073/pnas.0908533107>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing : Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <http://doi.org/10.1016/j.jml.2012.11.001>
- Boer, B. De. (2000). Self-organization in vowel systems. *Journal of Phonetics*, 28(4), 441–465.  
Retrieved from <http://www.sciencedirect.com/science/article/pii/S0095447000901256>
- Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1(2), 274–291. doi: 10.1111/j.1756-8765.2009.01019.x
- Brentari, D. (2007). Sign language phonology: issues of iconicity and universality. In E. Pizzuto, P. Pietrandrea, and R. Simone (Eds), *Verbal and Signed Languages, Comparing Structures, Constructs and Methodologies*, pp. 59-80.
- Brentari, D. (2017) Sign Language Phonology. In G. Gertz and P. Boudreault, (Eds.), *The Deaf Studies Encyclopedia*. Sage Publishers.
- Bush, R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, 68, 313-323.

- Buz, E., Tanenhaus, M.K., & Jaeger, T.F. 2016. Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*. doi: 10.1016/j.jml.2015.12.009
- Bybee, J. (1999). Usage-based phonology. *Functionalism and formalism in linguistics*, 1, 211-242.
- Bybee, J. (2003) Mechanisms of Change in Grammaticization: The Role of Frequency, in The Handbook of Historical Linguistics (eds B. D. Joseph and R. D. Janda), Blackwell Publishing Ltd, Oxford, UK. doi: 10.1002/9780470756393.ch19
- Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language* 82(4): 711-733.
- Caldwell, C. A., & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PLoS ONE*, 7(8). <http://doi.org/10.1371/journal.pone.0043807>
- Carrigan, E. (2016). An experimental investigation of the factors supporting the emergence of spatial agreement in Nicaraguan Sign Language. Unpublished doctoral dissertation: University of Connecticut. Retrieved from <http://digitalcommons.uconn.edu/dissertations/1299>
- Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, 110(28), 1 1278–11283. doi: 10.1073/pnas.1309518110

- Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences*, 112(7), 201418838. <http://doi.org/10.1073/pnas.1418838112>
- Dahl, O., Velupillai, V. (2013). The Perfect. In M. S. Dryer, & M. Haspelmath (Eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/68>, Accessed on 2017-08-01.)
- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Systematicity in Language. *Trends in Cognitive Sciences*, 19(10), 603–615. <http://doi.org/10.1016/j.tics.2015.07.013>
- Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26(5), 489–504. [http://doi.org/10.1016/0749-596X\(87\)90136-7](http://doi.org/10.1016/0749-596X(87)90136-7)
- Frishberg, N. (1975). Arbitrariness and Iconicity : Historical Change in American Sign Language. *Language*, 51(3), 696–719. <http://doi.org/10.2307/412894>
- Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, 29(5), 737–767. [http://doi.org/10.1207/s15516709cog0000\\_34](http://doi.org/10.1207/s15516709cog0000_34)
- Galantucci, B., Garrod, S., & Roberts, G. (2012). Experimental Semiotics. *Language and Linguistics Compass*, 6(8), 477–493. <http://doi.org/10.1002/lnc3.351>
- Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, 62(1), 35–51. <http://doi.org/10.1016/j.jml.2009.09.002>
- Gardner, M. (1970). Mathematical Games – The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, 223: 120–123.

- Gong, T., Baronchelli, A., Puglisi, A., Loreto, V. (2012). Exploring the Roles of Complex Networks in Linguistic Categorization. *Artificial Life*, 121, 107–121.
- Hall M., Richie R. and Coppola M. (2016). The impact of communicative network structure on the conventionalization of referring expressions in gesture. In S.G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher & T. Verhoef (Eds.) *The Evolution of Language: Proceedings of the 11th International Conference (EVLANG11)*. Available online: <http://evolang.org/neworleans/papers/134.html>
- Haspelmath, M. (1999). Optimality and diachronic adaptation. *Zeitschrift Für Sprachwissenschaft*, 18.2, 180-205.
- Heinz, J., & Idsardi, W. (2013). What Complexity Differences Reveal About Domains in Language. *Topics in Cognitive Science*, 5(1), 111–131. <http://doi.org/10.1111/tops.12000>
- Hockett, (1960) The Origin of Speech, *Scientific American* 203, 88–111 Reprinted in: Wang, William S-Y. (1982) *Human Communication: Language and Its Psychobiological Bases*, *Scientific American* pp. 4–12
- Jacobs, C. L., Yiu, L. K., Watson, D. G., & Dell, G. S. (2015). Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. *Journal of Memory and Language*, 84, 37–48. <http://doi.org/10.1016/j.jml.2015.05.004>
- Jaeger, T.F., & Buz, E. (2016.) Signal reduction and linguistic encoding. In E. M. Fernández & H. S. Cairns (Eds.), *Handbook of Psycholinguistics*. Wiley-Blackwell.
- Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, 67(3), 311–325. <http://doi.org/10.1016/j.jml.2012.07.002>

- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America*, 105(31), 10681–6.  
<http://doi.org/10.1073/pnas.0707835105>
- Sagi, E., Kaufmann, S., & Clark, B. (2011). Tracing semantic change with Latent Semantic Analysis. In Allen, Kathryn and Justyna Robinson (Eds.), Current Methods in Historical Semantics, pages 161-183. De Gruyter Mouton.
- Kocab, A., Pyers, J., & Senghas, A. (2014). Referential shift in Nicaraguan Sign Language: a transition from lexical to spatial devices. *Frontiers in Psychology*, 5, 1540.  
<http://doi.org/10.3389/fpsyg.2014.01540>
- Lam, T. Q., & Watson, D. G. (2014). Repetition reduction: Lexical repetition in the absence of referent repetition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(3), 829.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.
- Lewis, David, 1969. *Convention*, Cambridge: Harvard University Press.
- Magid, R. W., & Pyers, J. E. (2017). ‘I use it when I see it’: The role of development and experience in Deaf and hearing children ’ s understanding of iconic gesture.” *Cognition*, 162, 73–86. <http://doi.org/10.1016/j.cognition.2017.01.015>
- Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2), 313–318.  
<http://doi.org/10.1016/j.cognition.2012.09.010>

- Meir, I., Aronoff, M., Sandler, W. & Padden, C. (2010). Sign languages and compounding. In S. Scalise & I. Vogel (Eds.): *Compounding*. John Benjamins, 301-322.
- Morford, J. P., & Kegl, J. A. (2000). Gestural precursors to linguistic constructs: how input shapes the form of language. *Language and gesture*, 2, 358.
- Namboodiripad, S., Lenzen, D., Lepic, R., & Verhoef, T. (2016). Measuring conventionalization in the manual modality. *Journal of Language Evolution*, 1(2), 109–118.  
<http://doi.org/10.1093/jole/lzw005>
- Nowak, M., & Krakauer, D. (1999). The evolution of language. *Proceedings of the National Academy of Sciences of the United States of America*, 96(July), 8028–8033.  
<http://doi.org/10.1073/pnas.96.14.8028>
- Osugi, Y., Supalla, T., & Webb, R. (1999). The use of word elicitation to identify distinctive gestural systems on Amami Island. *Sign language & linguistics*, 2(1), 87–112. Retrieved from <http://cat.inist.fr/?aModele=afficheN&cpsidt=1554456>
- Pérez, A. (1990). Time in motion: Grammaticalization of the *be going to* construction in English. *La Trobe University Working Papers in Linguistics* 3, 49-64.
- Perlman M, Dale R, Lupyan G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, 2(8), 150152. <http://dx.doi.org/10.1098/rsos.150152>
- Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology*, 1, 227.  
<http://doi.org/10.3389/fpsyg.2010.00227>
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, 108(9), 3526–9. <http://doi.org/10.1073/pnas.1012551108>

- Pierrehumbert, J. B., Stonedahl, F., & Daland, R. (2014). A model of grassroots changes in linguistic systems, 30. Retrieved from <http://arxiv.org/abs/1408.1985>
- Pinker, S. & R. Jackendoff. (2005). The faculty of language: What's special about it? *Cognition*, 95, 201–236.
- Quine, W. V. O. (1960). *Word and Object: An Inquiry into the Linguistic Mechanisms of Objective Reference*. Cambridge, MA: MIT Press.
- Rescorla, M. (2015). Convention. In E.N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), URL = [<http://plato.stanford.edu/archives/sum2015/entries/convention/>](http://plato.stanford.edu/archives/sum2015/entries/convention/).
- Reynolds, C.W. (1987) Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, 21, 25–34.
- Richie, R. (2016). Functionalism in the lexicon: Where is it, and how did it get there? In G. Jarema, G. Libben, & V. Kuperman (Eds.), Thematic Issue of *The Mental Lexicon: New Questions for the Next Decade*, 11(3), 429–466. Amsterdam: John Benjamins. doi 10.1075/ml.11.3.05ric.
- Richie, R., Yang, C., & Coppola, M. (2014). Modeling the emergence of lexicons in homesign systems. *Topics in Cognitive Science*, 6, 183-195.
- Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, 141, 52–66. <http://doi.org/10.1016/j.cognition.2015.04.001>
- Roche, J. M., Dale, R., & Kreuz, R. J. (2010). The resolution of ambiguity during conversation: More than mere mimicry? In R. Camtrabone & S. Ohlsson (Eds.), *Proceedings of the 32<sup>nd</sup>*

- Annual Meeting of the Cognitive Science Society (pp. 206-211). Austin, TX: Cognitive Science Society.
- Sandler, W., Aronoff, M., Meir, I., & Padden, C. (2011). The gradual emergence of phonological form in a new Language. *Natural Language and Linguistic Theory*, 29(2), 503–543.  
<http://doi.org/10.1007/s11049-011-9128-2>
- Sandler, Wendy & Lillo-Martin, Diane (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.
- Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive psychology*, 21(2), 211-232.
- Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, 41(3-4), 249-263. doi: 10.1016/j.wocn.2013.03.007
- Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary theory and the ultimate–proximate distinction in the human behavioral sciences. *Perspectives on Psychological Science*, 6(1), 38-47.
- Senghas, A., & Coppola, M. (2001). Children creating language: how Nicaraguan sign language acquired a spatial grammar. *Psychological Science : A Journal of the American Psychological Society / APS*, 12(4), 323–328. <http://doi.org/10.1111/1467-9280.00359>
- Senghas, A., Kita, S., Ozyürek, A. (2004). Children Creating Core Properties of Language: Evidence from an Emerging Sign Language in Nicaragua. *Science*, 305(5691), 1779–1782.  
<http://doi.org/10.1126/science.1100199>.
- Smith, K. (2002). The cultural evolution of communication in a population. *Connection Science*, 14(1), 65–84.

- Steels, L., & Loetzsch, M. (2012). The grounded naming game. In L. Steels (Ed.), *Experiments in cultural language evolution* (pp. 41–59). Amsterdam: John Benjamins.
- Stent, A. J., Hu\_man, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication*, 50 (3), 163-178. doi: 10.1016/j.specom.2007.07.005
- Spike, M., Stadler, K., Kirby, S., & Smith, K. (2016). Minimal Requirements for the Emergence of Learned Signaling. *Cognitive Science*, 1–36. <http://doi.org/10.1111/cogs.12351>
- Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, 4, 357-380.
- Verhoef, T., Kirby, S. & de Boer, B. (2015). Iconicity and the emergence of combinatorial structure in language. *Cognitive Science*. doi:10.1111/cogs.12326
- Wedel, A. (2012). Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition*, 4(4), 319–355. <http://doi.org/10.1515/langcog-2012-0018>
- World Federation of the Deaf (2016). *Sign Language - WFD*. Retrieved from <https://wfdeaf.org/human-rights/crpd/sign-language/>
- Zuidema, W., & Boer, B. De. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, 37, 125–144. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0095447008000624>

## **Appendix A. Experimental items**

avocado  
baseball cap  
beans  
boy  
cabbage  
cloud  
cow  
cowboy hat  
dog  
girl  
goat  
horse  
lake  
lime  
man  
mountain  
old woman  
orange  
pig  
policeman  
rice  
sheep  
soldier  
truck  
woman