UCLA
UCLA Electronic Theses and Dissertations

Title

Emerging Communication Systems: Interaction for Language Evolution and Transmission

Permalink

https://escholarship.org/uc/item/6qg6p4pg

Author

Micklos, Ashley Michelle

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Emerging Communication Systems:

Interaction for Language Evolution and Transmission

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Applied Linguistics

by

Ashley Michelle Micklos

2015

ABSTRACT OF THE DISSERTATION

Emerging Communication Systems:

Interaction for Language Evolution and Transmission

by

Ashley Michelle Micklos

Doctor of Philosophy in Applied Linguistics

University of California, Los Angeles, 2015

Professor John H. Schumann, Chair

Interaction is a significant and dynamic aspect of human language use, however, investigations into the emergence and evolution of language do not adequately consider how interaction facilitates such processes. In this dissertation, interaction is considered both broadly- as in among the "causes" behind language emergence- and specifically- as between co-participants in a language use context. First, the complex adaptive systems approach is applied to the many theories of language origins and evolution, proposing that multi-causality brought about through interacting forces can lead to language emergence in humans. Looking at how children learning language interact with caregivers, the ability to pull together resources to make meaning, even without full language, becomes clear. This ability from natural language use and learning is what informs the experimental investigation of interaction's affect on language emergence and

evolution. In the lab, we have used the iterated learning paradigm (Kirby, Cornish, & Smith, 2008) but adapted it to face-to-face interaction involving a gradual turn-over of participants to simulate transmission over generations. Using silent gesture (Goldin-Meadow et al, 2008; Schouwstra, 2012), participants in the first experiment communicated with and matched gestures to a selection of target images involving a ball moving in a specific manner and path. Over generations, gesture time and diversity decreased (participants' gestures became more aligned). Moreover, lineage-specific eye gaze patterns evolved, which, when deviated from, indicated a need for repair on the gestured form. These repairs, often in the form of clarifications, made elements of the gesture more salient, leading to their fixation in the system. A second study used the same interactive paradigm as the first, but incorporated a condition in which repair could be performed in a third-turn, as it typical of natural conversation. Having to disambiguate easily confusable noun-verb pairs using silent gesture, participants were allowed a "do-over" repair turn or not; both conditions developed a systematic noun marking system, though repair condition chains did so at a quicker rate. More importantly, the increased interactivity, namely via negotiation and repair, drove systematicity more rapidly than previous non-interaction studies.

The dissertation of Ashley Michelle Micklos is approved.

Gregory A. Bryant

Marjorie Harness Goodwin

Charles Goodwin

John H. Schumann, Committee Chair

University of California, Los Angeles

2015

Dedicated to my Mom and Dad, for their support, guidance, and love.

Table of Contents

List of Tables

<p style="text-align:center">List of Figures</p>

List of Images

Acknowledgements

I am greatly indebted to my mentors, advisors, and committee members for their guidance, feedback, advice, and general good-naturedness. I would like to acknowledge Erica Cartmill for her feedback on lab experiments, but for also being a role model for women in our field. To Clark Barrett, signing member of my qualifying paper, thank you for pushing me to think critically, and convincing me to do experimental work. To Candy Goodwin, thank you for your positivity- you kept me motivated throughout this process and I always felt great encouragement from your feedback. I owe a lot to Greg Bryant, who planted the idea to visit a field-leading university for research training. You have given me many opportunities that have benefited my study, persuaded me to use statistics, and, most significantly, allowed me the use for your lab and participant pool for my research. I am very appreciative of Chuck Goodwin, professor and committee member for my graduate studies. I believe the training I have received from you, Chuck, is invaluable, and I am honored to have been able to share my work with you over the years. And to my advisor, John Schumann, you set me on the path of language evolution research, and I am so thankful that you encouraged me to pursue it in light of the challenges facing interdisciplinary work. I think words fail to convey the respect I have for you as a mentor and applied linguist, and I will be forever grateful for your comments, advice, and humor.

Much appreciation goes to Susan Plann, for nominating me for the Cota-Robles Fellowship, which provided funding for my graduate studies, including tuition, research, and travel. Also, many thanks to the Applied Linguistics department, for their financial and academic support.

I was fortunate to visit the University of Edinburgh's Language Evolution and Computation research unit- an experience that enabled much of my research and gave me a wonderful life experience. I would like to thank Monica Tamariz, Marieke Schouwstra, and Kenny Smith for their suggestions, guidance, and feedback while devising laboratory experiments and presenting data. I am truly grateful to have been under the supervision of Simon Kirby, who presented me with an incredible opportunity to embark on a new line of research, and supported it with training and feedback. Most importantly, I must give credit to my collaborator, friend and partner-in-karaoke, Yasamin Motamedi, who will be a co-author on the research presented in Chapter 3 (in prep). We co-devised the interactive iterated learning model and implemented its first experimental trial. She tirelessly coded handshapes, ran statistical analyses on entropy, and insightfully interpreted findings. I would also like to recognize the entire LEC cohort (2013-2014), who made me feel welcome in a new research environment and country- your friendships mean the world to me!

Finally, friends and family have been my support system during my graduate studies. I would like to acknowledge my fellow UCLA Applied Linguists and friends for commiserating, bending an ear, happy hour, and the encouragement. Also, a thank you to my friends in San Diego, who I've known and loved for many years- thank you for making my life decision to earn a PhD seem just as impressive as all of your decisions! Above all, I am so thankful to have had the love and support from my family, most especially my parents, as I have delayed a "normal" life to pursue my passion. My parents provided me so many more opportunities than they were given and instilled in me a work ethic that has been demonstrated here. I love you both very much.

ASHLEY MICHELLE MICKLOS

EDUCATION

University of California, Los Angeles – June 2010
Master of Arts, Applied Linguistics and TESL – GPA: 3.93
Thesis: Peer Corrective Feedback in the Second Language Classroom
Chair: Professor Susan Plann

University of Nevada, Las Vegas – December 2006
Bachelor of Arts, Spanish – GPA: 3.81
University Honors- *Magna Cum Laude*

RESEARCH EXPERIENCE

2014 - 2015   PhD Dissertation Experiment, Primary Investigator

University of California, Los Angeles

- Establish experiment methodology; create stimuli and consent materials
- Pilot experiment for timing, task difficulty, and logistical measures
- Code and analyze data, both qualitative and quantitive measures

2013 - 2014   Language Evolution and Cognition Research Unit, Visiting Postgraduate Student

University of Edinburgh

- Continue previously piloted work into a dyadic setting
- Co-create experiment design, interface, and materials
- Run experiment internally
- Code and analyze interaction features, timing, and eye gaze
- Present results in weekly lab meetings and unit talk series

PUBLICATIONS

Micklos, A. (2014). The nature of language in interaction. In E.A. Cartmill, S. Roberts, H. Lyn, and H. Cornish (Eds.), *The Evolution of Language: Proceedings of the 10th International Conference (EVOLANG10)*, World Scientific: New Jersey.

CONFERENCE PRESENTATIONS AND INVITED TALKS

Micklos, A. (2015). *Iconicity in emerging silent gesture communication: Influences from repair strategies.* Conference Presentation at the 10th International Symposium on Iconicity in Language and Literature, University of Tubingen, Germany.

Micklos, A. (2015). *Interaction's role in emerging communication systems and their conventionalization: Repair as a means for the fixation of form-meaning matches.* Invited talk at the Center for Research in Language, University of California, San Diego.

Micklos, A. (2014). *Interaction's role in emerging communication systems and their conventionalization: Eye gaze, turn-taking, and repair.* Invited talk at theLanguage Evolution and Cognition Lab, University of Edinburgh.

Micklos, A. (2014) *The nature of language in interaction.* Poster Presentation at the 10th Evolution of Language Biennial Conference, University of Vienna.

Micklos, A. & Motamedi, Y. (2014). *You can't hand**le the truth: A gestural communication study.* Language Evolution and Cognition Lab Presentation, University of Edinburgh.

Micklos, A. (2013). *Human co-operative action: How much is "human?"* Poster Presentation at the 25th Human and Behavior Evolution Society Annual Conference, Miami, FL.

Micklos, A. (2012). *De-fossilization: Possibilities from learner characteristics.* Conference Presentation at the Linguistics Symposium, Purdue University.

Micklos, A. (2010). *Peer-to-peer corrective feedback in the foreign language classroom.* Conference Presentation at the 5th UC Consortium Conference on Theoretical and Pedagogical Perspectives, UC San Diego.

Micklos, A. (2010). *Peer corrective feedback: New categories.* Conference Presentation at the Second Language Acquisition and Teaching (SLAT) Roundtable, University of Arizona.

HONORS AND AWARDS

2010 - 2014   Cota-Robles Fellowship Recipient, *UC Los Angeles*
2009 - 2010   Continuing Graduate Student Fellowship, *UC Los Angeles*

COURSES TAUGHT

Summer 2012-2015   Applied Linguistics 40W: Introduction to Language and Gender, *UCLA*
Spring 2013, 2015   Applied Linguistics 102W: The Nature of Learning, *UCLA*
Summer 2014   ESL 32: Conversation for Academic Interactions, *UCLA*
Summer 2012/2013   ESL 97A: American Culture Through Film, *UCLA*
Summer 2013   ESL 34: Public Speaking, *UCLA*
Winter 2012   Applied Linguistics 101W: Language Learning and Teaching, *UCLA*
Fall 2011   Applied Linguistics 10W: Language in Action, *UCLA*
2008 - 2009   Spanish 1-3: First-year Spanish, *UCLA*

PROFESSIONAL ACTIVITIES

2012   Applied Linguistics Conference Committee Co-Chair, *UC Los Angeles*
2011   Applied Linguistics Conference Committee Member, *UC Los Angeles*

CHAPTER 1

AN INTRODUCTION TO LANGUAGE EVOLUTION AND EMERGENCE

We use language to accomplish a great many things: make transactions and business deals, build relationships between family and friends, teach children about the world, and establish communities of practice based on common goals and actions. Our use of language includes more than just words strung together in meaningful, symbolic, and abstract ways, but also incorporates gestures, body movements, eye gaze, prosody, and other paralinguistic variables. Language is not used in isolation, in terms of context or form. We use language with the environment and in the environment. That is, we integrate the resources available to us when we use language, and then use those aspects jointly to accomplish tasks, goals, actions, and other interactions. Language is the action we perform to carry out further actions, especially those involving co-participants. Our tremendous sociality as a species requires cooperative behavior-what better way to get things done than with a communication system, be it vocal or gestural.

The focus of this research is on the nature of language in interaction, from an evolutionary perspective. While most investigations into the evolution of language are concerned with syntactic structures and symbolic meaning, this research centers on a more fundamental meaning of language: that of communication and language use. This focus is on the pragmatics and function of language, rather than its structure. Specifically, the studies here are aimed at examining how features of interaction facilitate the evolution of communication systems, and how they are transmitted with the system. The tradition of conversation analysis views talk as action in interaction and emphasizes what talk *does* (Heritage & Clayman, 2012). This tradition will inform the current research on language evolution, investigating the parallels between

language as action today and how it may have been carried out in an evolutionary environment. As Kendon (2009) proposes, if we combine vocal and gestural communicative forms to make meaning and complete action in modern times, why should this not also be true for our ancestors pioneering communicative techniques? Language as a complex adaptive system must rely on a multitude of resources, materials, and modes for "use" (understandability, transmission, etc)- these should be considered in theory of language origins.

LANGUAGE EVOLUTION THEORIES

Over the past few decades language emergence and evolution has garnered much attention in the field of linguistics. Since reigniting interest in language origins many researchers have speculated about the basis for language in our earliest ancestors. Even researchers from varied backgrounds, including genetics, paleo-anthropology, and neuroscience, have weighed in on the debate, yet there still remain epistemological issues in proposing theories about language origins. The purpose of this chapter is to examine evolutionary theories of language by providing a critical literature review. Further, the goal is to propose an inclusive approach to addressing these issues.

First, it will be helpful to distinguish between language emergence and language evolution. Many theorists often fail to discriminate between these terms, and as such their hypotheses may confound them, which could be problematic both epistemically and empirically. When using it in terms of language, "evolution" becomes problematic as it could bear two meanings: the evolution for language, and the evolution of language (i.e. specific languages). The perspective of this chapter takes on the former, though many researchers in the field propose theories of the evolution of language from protolanguage to complex language. (Also, linguists

already have a subfield for tracing language change over time, called historical linguistics). The term "emergence" suggests the point of origin of language; extremists may consider the discussion of "language emergence" as taking language as an emergent, or "pop-up," event that sprang into being. However, emergence and evolution may share a common ground when placed in the realm of language origins. When addressing the "how" and "why" of vocal communication, its emergence may be as gradual as its evolution. That is, how language emerged and what evolved for language in humans may in fact be asking the same question. In this case, we might use the terms interchangeably. It should be noted that I will use the terms "evolution" and "emergence" as originally used by the authors (though, many do not distinguish between or define them) in the accounts of the literature given below.

I will first outline the problems that arise with these theories, including the problem of considering only a single cause for language emergence. As will be discussed later, human evolutionary history is not made up of linear events, but rather consists of multiple levels of change. With brains, culture, sociality, and cognitive abilities all becoming more "human-like" simultaneously, it seems impossible to pin down one of these as the root of human language. Additionally, any theory is only as good as its backing. Therefore, evolutionary accounts for language must not only provide the "how" of language emergence/evolution, but the "why." While some evolutionary theories of language do propose both the how and why of language origins, important questions remain unaddressed. That is, the reader is still asking, "where did that come from?" Evolutionary theorists must consider the support for their theories, going back as far as the evidence will allow.

In the following section I will provide more depth on the issues mentioned briefly above, elaborating on how each one dampens the strength of language evolution theories. Then, I will detail the most current and prominent language evolution theories, noting the how the theorists tackle the issues mentioned in section two (or perhaps fail to). Afterward, I will review Lee et al's (2009) account of language as a complex adaptive system, positing how evolutionary theories of language would fit into the model. Finally, I will argue the need to consider language evolution from multiple sources, including but perhaps not limited to the current theories discussed here.

CRITICISMS OF LANGAUGE EVOLUTION THEORIES

Language evolution and emergence theory-building is an inherently challenging task. Little evidence for language exists in fossil records of our ancestors. Little is known about the role genes may play in language emergence. And, pinpointing a direct cause for language has been a futile effort. With these challenges, and likely many more, it is only human to "just do what you can" when it comes to theorizing about language origins. Nonetheless, we must strive to theorize in the most complete and sound manner possible. Here I propose two critical issues that arise in language evolution theory, which, if addressed, could hopefully result in greater consensus among researchers, scholars, and theorists.

The Problem with Linear/Singular Causality

Scholars have to write. Even more, scholars need to propose theories that are at least minimally distinct from one another, resulting in at least slight variation and contention. A researcher's fame (whether it be good or bad) derives from his theory, and he must therefore make it stand out- his claim to fame is the opinion he proposes as "the one." Unfortunately, it

does not seem likely that language has a singular cause. The social and cognitive milieu in our environment of evolutionary adaptation was ever increasing in complexity and variety. Natural selection could also have a hand in this development as well, affecting cognitive abilities and anatomical structures. Settling on one cause or one pathway to language is unreasonable when considering the wealth of features we employ just to use it today. Though, the one cause-theory does not limit action in the debate, as any other scholar could simply suggest one of the multiple other causes that likely contributed to language emergence and evolution. This however should not be the impetus for theory development; rather we should strive to propose plausible theories that consider the multi-causality of language origins. This point will be further discussed below.

Giving the Full Story: "Where did that come from?"

Given the difficulties in obtaining evidence for language origins, theorists often fail to address all aspects of the evolutionary story: why and how did it emerge/evolve? Why did it develop at all, or in the way it did? The "how" aspect, on the other hand, is crucial to any evolutionary theory and ties in closely with the "why." To account for the "how" of language emergence/evolution would require us to consider what came before language. What was in our evolutionary past that brought about this phenomenon? As we will see, some researchers suggest social motives, increased cognitive abilities, or even directly by natural selection for language itself.

CRITICAL LITERATURE REVIEW: LANGUAGE EVOLUTION THEORIES

Theories on the origins and evolution of language can be divided into (at least) three sub-categories of "theory type." These are socially-driven, cognition-driven, and biologically-driven theories of language evolution. There is no doubt still variance within each category, and overlap

between them, but this division distinguishes the main types of causation offered in these theories.

SOCIALLY-DRIVEN LANGAUGE EVOLUTION

Evolutionary theories of language that are socially-driven posit that language was the result of increasingly complex sociality in early human societies. Some researchers claim language was the crux of social interactions of our ancestors, while others believe social systems provided an environment in which language could emerge.

Dor and Jablonka (2010) suggest greater behavioral plasticity in general learning capacities led to the social emergence and evolution of language. That is, rather than a language gene being selected for, learning abilities already in place were selected to aid language emergence. The authors suggest that the social evolution of language can be attributed to bigger brains (and associative learning), better memory, better skills of social engagement, sophisticated versions of theory of mind, and better imitation skills. Once languages were socially invented, they argue, the structures that remained were adapted to the brain. For instance, the structures of language that best suited our learning capacities. While this theory is essentially social, genes may play a role in the emergence of language, though perhaps not in its evolution per se. Dor and Jablonka posit language emerged from a co-evolutionary spiral in which innovations lead to increased theory of mind that then brings about the ability to learn. In this co-evolutionary spiral, language not only adapted to human brain, but the brain also adapted to language. The authors identify the learnability of languages (e.g. not too complex structures) as an adaptation to the brain, while the brain adapted in its ability to perceive communicative intentions and linguistic input.

In another socially-based view, Worden (1998) views language as emerging from social intelligence, a feature passed down through our primate lineage (or, "pancestors," a term coined by Mark Muller). Worden implicates social intelligence and theory of mind as the co-opted pre-requisite abilities for language. His "how" account of language evolution is based on evolutionary constraints, namely the "evolutionary speed limit" in which the computational faculties underlying language must be based on pre-existing capacities. In this case, social intelligence (from primates) fits the bill, as Worden suggests no further cognitive faculties needed to develop (beyond this point) for language to emerge. Furthermore, within this perspective, language is used for social communication and thus must be linked to social intelligence (especially theory of mind). Worden's account of co-optating for language is supported by evidence from neurophysiology that finds common evolutionary origins in the brain for language and primate social intelligence in the ventral pre-frontal cortex (VPC), and Broca's area frontal-most part for the semantic and syntactic aspects of language. The area of Broca's overlaps the VPC. Though Worden does not present a view on "when" or "why" language emerged, he does view Dunbar (1998) and Power's (1998) social brain/gossip hypotheses as compatible with his own "how" account.

Dunbar's (1998) social brain hypothesis focuses on the group bonding that took place in early hominid groups, wherein the brain (and eventually language) evolved to solve social problems not ecological ones. He proposes highly structured, large social groups put selective pressure on brain growth. This is illustrated well in a study showing a dramatic growth in neo-cortex size positively correlated with increased group size for both primates and humans. Dunbar asserts that unlike primates, though, early humans could not devote as much time to social

behaviors, namely grooming, thus language emerged as a solution to this social problem (that is, maintaining social relations when time to do so is not provided for in the environment). Language is an effective solution because it allows better management of time (you can "groom" more than one individual at a time) and it helps build relationships without physical contact through the direct exchange of information. To this end, Dunbar claims the principle function of language is the exchange of social information, or gossip, to facilitate bonding as a consequence of larger groups. The difference between Dunbar's theory and other theories is purely a matter of the goal of such social interactions; for Dunbar, the goal is the exchange of social information, while Dessalles (2000), for instance, suggests status attainment is the fundamental purpose for the exchange of information.

Power (1998), however, questions the "gossip" hypothesis by problematizing the reliability of cheap, or easy to fake, signals (i.e. words) for the exchange of information, particularly when cheater detection would need to be in place. Power attributes the propensity to fake signals to Machiavellian intelligence; she assumes, however, that Machiavellian intelligence only accounts for tactical deception in individuals, and not cooperation. However, since an individual can lie with words, we could question the reliability of cheap signals if we believe language emerged for political coalition building (Dessalles, 2000). Though, if the goal of language were to establish and maintain social bonds, ones founded on fake signals would not benefit in the long-term. Furthermore, it seems likely that a "free-rider" detection system could have been in place for this very reason. Current research in evolutionary psychology demonstrates how free-riding detection appears to be its own mechanism, and that it works quite well over a variety of contexts (Denton et al, 2012). In fact, Dessalles' (2000) cooperation theory

of language emergence claims we can detect deception as listeners by using logic, which could have also evolved to make lying difficult and detectable. He proposes that early language users would be in a constant process of gathering new information and checking its consistency among other information in order to establish friendships with reliable individuals.

To return to Dunbar's theory, he also proposes an origin of language, though not with much precision. He believes the need for more effective social bonding (via language, that is) arose between 500,000-250,000 years ago, positing that *Homo erectus* may have had language, but that early *Homo sapiens* definitely did. Additionally, he proposes an evolution of language in which it increased in complexity rather than suddenly appearing without any precursors. A simple explanation that he offers is that language started with simple vocalizations and then developed into a more symbolic form. In sum, Dunbar's theory provides an account of the why and how of language evolution and emergence (without confounding the two), and even proposes a "when" account for language origins, albeit a vague one. While Dunbar's theory is consonant with other social grooming theories, it still only implicates this one social cause for language emergence. We might, though, hypothesize that Dunbar's theory also includes hints of cognitive ability, given the relationship between increased sociality in larger groups and brain size.

Gesture is yet another area in which researchers look to divine the origin of language in a social context. Donald (1998) posits mimesis was the first element for social cohesion and communication, as well as culture. This manual interaction eventually evolved into spoken language; Corballis (2010) suggests a similar story. He reconciles two prior theories, those of language by natural selection and language as a pop-up phenomenon, by proposing humans were

pre-adapted for gestural communication since apes also have this ability (thus it arose before the ancestral split), but that a switch from visual to aural language occurred gradually. Evidence for this could be the homologous brain region in which we find mirror neurons in primates and humans. He claims that gestures were common among primates and early humans, as bipedalism allowed for the use of hands for communication. However, as hand tools became a necessary survival resource for early humans (and hands became occupied for the craft), spoken language began to emerge, and with it, more sociality and complex communication. Vocal language had many benefits to its new users, including the use of less energy, the freedom of hands for using tools, communication over longer distances, and increased memory processing. Corballis' account appears to satisfy the "whole" story requirement, and while he implicates gesture as the impetus for language emergence, he includes other aspects of human evolutionary history as corollary factors, including tool use (which we may presume implies some cognitive abilities), bipedalism, and increased sociality.

Kendon (2009) notes, however, gesture should not be considered the precursor to vocal language, given its use in conjunction with vocal communication. That is, he argues language in modern use is not an isolated phenomenon, but instead is used with gesture involving the hands, body, and face. Rather than conceptualizing a switch from gestural to vocal communication, Kendon asserts a polymodalic communication action for forming symbolic referents. The author also concedes that our understanding of how our ancestors came to recognize (and differentiate) symbolic and referential action is not complete. In a commentary on Tomasello (2008), Kendon (2009) suggests that the investigation of language origins consider the "nature of the cognitive

and social matrix which language surely requires" (p. 370). Taking the approach Kendon advises would be a step in a more inclusive direction in language origins theory.

Tomasello et al (2005) champion a more complex version of theory of mind, a socio-cognitive skill found in primates and autistic children but not to the same degree of intention reading as in normally developing human children: shared intentionality. While not explicitly a theory of language evolution, this perspective on higher cognitive abilities in humans does lend some insightful information to the debate. Tomasello and his colleagues propose two hypotheses about the development of such increased abilities in sharing and reading the intentions, beliefs, and motivations of others: an ontogenetic one and a phylogenic one. The latter argues humans evolved through not just competitive motivations, but collaborative ones that involved shared goals and joint attentions. The individuals who were most collaborative would have a selective advantage and perpetuate the skill. Evolving more "human-like" collaborative skills requires a number of processes, including more within-group tolerance, motivation to share emotions with others (that is, to affiliate), and imitation. This hypothesis supplements Machiavellian theories of cognitive evolution, albeit with a cultural emphasis. In sum, human cooperative communication is an adaptation for collaborative activity and cultural life (Tomasello, 2008).

The ontogenetic hypothesis (Tomasello et al, 2005) states human infants are able to read intentions well and are motivated to share their emotions/internal states with others around age one. The seemingly rapid development of these abilities sets the path for the child to "participate in collaborative cultural practices" (p. 688). Again, though this is not an explicit theory of language emergence or evolution, these findings would be logical to include among such theories because they demonstrate a cognitive and cultural distinction between humans and our closest

relative. He argues that the socio-cognitive infrastructure of human cooperative communication is quite complex, as it comprises the skills of shared intentionality (Tomasello, 2008). Tomasello and his colleagues additionally consider the phylogenic and ontogenetic development for shared intentionality, both of which may have been important for language evolution.

Hurford (2007) takes a supportive stance towards Tomasello's (2005) argument of shared intentionally as a means to get to (spoken) language, and he also concedes that many factors contributed to language evolution, specifically to a language-ready species. Hurford notes that apes too have mental representations and are in fact pre-linguistic. This suggests that man's linguistic abilities started as far back as the common ancestor of apes and humans. However, as Hurford argues, humans took an extraordinarily different path that required communicative cooperation (possibly arising from kin and/or cultural group selection), signal development, and trust (fostered by the pro-social hormone oxytocin).

Consistent with multiple causality, Chater and Christiansen (2010) argue that language evolution was forced by cultural evolution, with many constraints acting upon the shaping of language including perceptuo-motor, cognitive (e.g. limitations on learning), thought, and pragmatic factors. These constraints act in conjunction with one another so patterns of language may arise. This theory seems quite congruent with the Interactional Instinct's (to be discussed later) case for language as a complex adaptive system, in which patterns emerge from chaos- the multiple factors at play in the shaping of language. The authors deny that language could have evolved or emerged from innately specified modules in the brain, rather they posit language is adapted to the brain and transmitted culturally. In fact, they believe the biological structure was set prior to language, and cultural forms (e.g. language) adapt to fit the biological structure

12

through variation, propagation, and elimination. Further, cultural forms will also be modified by functional constraints (consider the modification of a tool for functional factors). In addition, Chater and Christiansen offer a prospective on language acquisition, claiming it relies on "acquiring the ability to coordinate with others" (p. 1137) and aiming to be like others. This is labeled C-induction (culture induction), an ability much easier than understanding the natural world (N-induction) because it is easier to learn through social cues and does not rely on the true/ false dichotomy, the authors assert. Since C-induction and language depend greatly on culture, and culture evolves rapidly, they are moving-targets for biological adaptation[1]. While Chater and Christiansen outline a detailed account of the constraints on shaping language to the brain and how language is transmitted culturally, they do not account for how these factors were assigned to language shaping initially. That is, the question of why these features were co-opted for language evolution remains unanswered.

The social causes for language emergence are many, and they often seem to overlap in their contribution (many theorists will posit more than one social factor behind language emergence). Figure 1.1[1] (below) shows this overlap, depicting the social causes for language detailed above, but also showing their correspondence with each other. Social abilities and phenomena hardly happen in a vacuum, but rather build upon previous social skills/abilities to form more complex ones- as can be seen with primate gestures (which with the increasing tool innovation gave way to mimesis and eventually advanced imitation abilities).
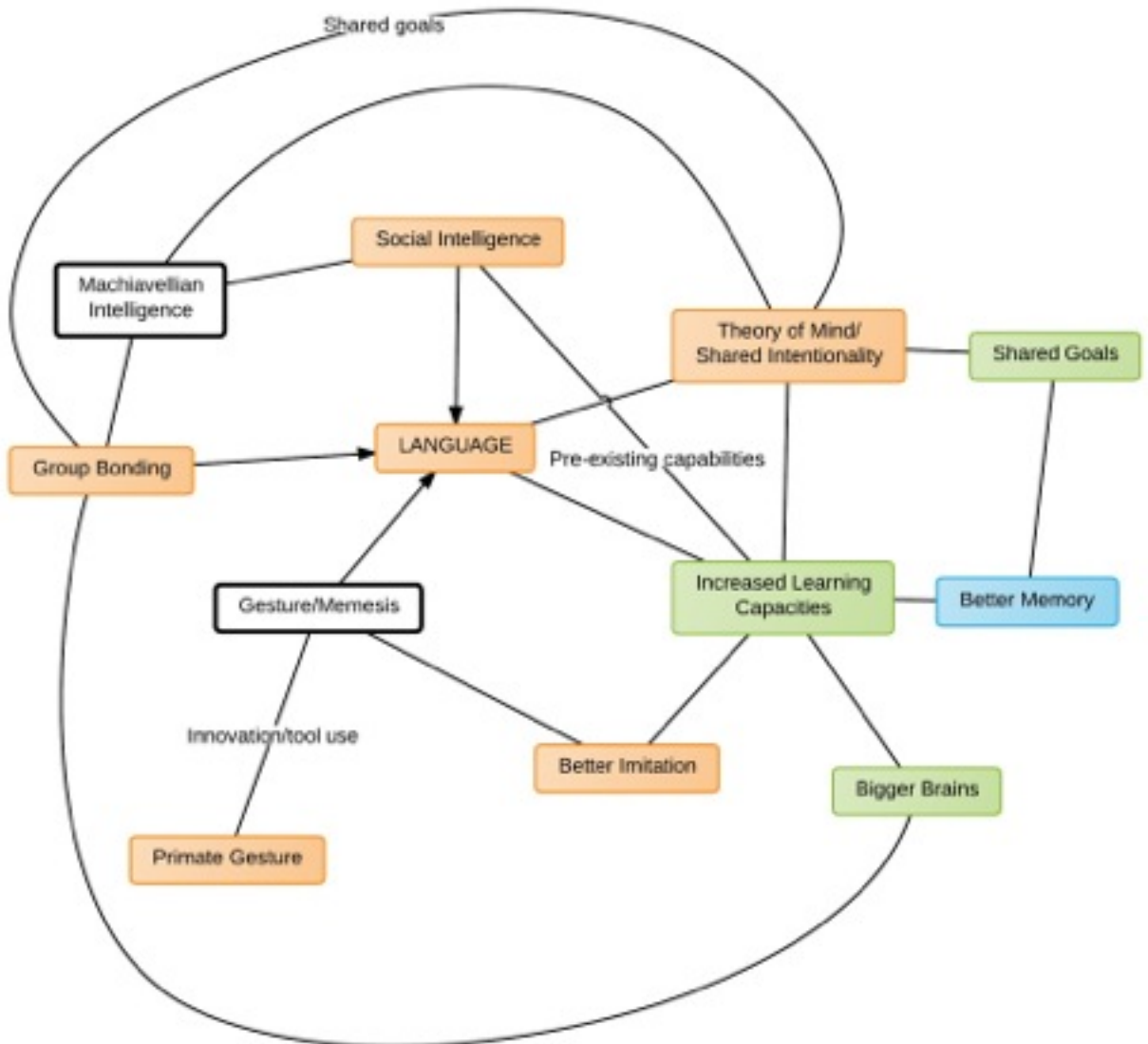
Figure 1.1: Social Causes for Language

Socially-based language evolution is in itself quite dynamic, but we shall see how it too interacts with cognitive and even biological factors involved in the emergence of language.

COGNITIVELY-DRIVEN LANGUAGE EVOLUTION

While separating cognitive and social causes for language origins can be challenging since much overlap exists between them, some researchers propose decidedly cognitive based theories for human language evolution.

Gärdenfors and Osvath (2010) argue prospective cognition, the ability to plan future goals, was the precursor to language among our ancestors, and not simply social interaction. Evidence that shows early hominids habitually carried tools over distances (supposedly to be used for a future goal) and divided labor (cooperation for shared goals) support the notion that they did have the ability to foresee future events and plan goals for them. According to the authors, "the cooperation about future goals requires that the inner worlds of the individuals be coordinated" (p.110). To communicate an individual inner world, Gärdenfors and Osvath propose, would require theory of mind and symbolic communication. They refer to Deacon's premise of symbolic communication, the first of such being a "marriage" contract, as an advanced form of cooperation for future goals. All in all, Gärdenfors and Osvath suggest prospection and symbolic communication co-evolved, leading to language in early humans. Their position places causation in the cognitive domain, though these capacities would seemingly involve increased sociality as well. The authors, however, do not consider this point in their otherwise plausible theory.

In contrast, Tattersall (2010) takes an exaptationist approach, positing symbolic cognitive abilities (in a symbol-ready brain) were co-opted for language at the onset of cultural innovation. Specifically, he asserts neural substrates for symbolic cognitive abilities arose as a by-product of physical re-organization and lay dormant until the onset of cultural innovation. Since form precedes function, he argues structures for speech were already in place before language emerged. Tattersall also claims the brain is symbol-ready and was co-opted for language in an emergent event. To pinpoint the time of emergence, Tattersall looks to the fossil record for evidence of cultural innovation. He determines that *Homo hiedelbergensis*, though having stone

tools, fire, and shelters, did not engage in symbolic activity and therefore likely did not have structured language (though he does allow vocalizations as a possibility). Neanderthals, he claims, also do not show much evidence of a symbolic consciousness despite their larger brains and increased care for group members. Many authors (Beaken, 2011; Hawks, 2011) would disagree on this point, claiming that Neanderthal burials are suggestive of cultural innovation and symbolic thought. Tattersall, then, leaves the only option for the emergence of language with Cro-Magnon, or *Homo sapiens*, claiming that findings of art, song, music, and carvings are evidence for symbolism. As an anthropologist without expertise in language per se, Tattersall admits his own lack in providing a how or why story to language evolution. Though, his study of human evolution, he believes, gives him insight into when certain cognitive abilities required for language emerged. Not all researchers would agree with such a late emergence of language (especially a simple one) in the *Homo* line, but perspectives on cultural intelligence and innovation may well be subjective.

Unlike many of his colleagues, Ulbaek (1998) looks further back into our evolutionary history for the link to language. He proposes language evolved in a continuous line of development from animal cognition, that is, cognitive systems that had long since been in place. Ulbaek elaborates on the cognitive abilities many animals (namely, apes) possess without language, including tool use/making, cognitive maps, learning through imitation, social knowledge, theory of mind, and deception. These findings assume a higher intelligence of many animals, from which our human intelligence evolved. From this perspective, the function of language is to communicate thoughts, which may not have been advantageous for all species. The costs associated with communicating thoughts through vocal language are quite high and

include more brain tissue, re-organization of the brain, and changes in the respiratory system. The benefit is that it enables cooperation, but by giving away information (not necessarily an evolutionary stable strategy, see Dunbar, Power, and Dessalles' arguments above). So, to enhance one's own fitness in a situation such as this, Ulbaek proposes language evolved with altruistic traits.

As such, Ulbaek does not completely discount the social purpose of language, adding that special social conditions, not superior intelligence, would be required for animal cognitive systems to evolve for language in *Homo*. One such condition is reciprocal altruism in which one gains fitness by sharing with or helping others. This condition inherently requires that free-riders be detected as well. Other requirements for reciprocal altruism include long lifespan, low dispersal rate, living in small mutually dependent groups, and a long period of parental care- all seemingly present in the *Homo* lineage. Though this view of language genesis is distinct from other researchers, Ulbaek does concede some level of interplay between the cognitive and social aspects of early hominid life that caused language to emerge. While Ulbaek provides the how of language emergence, the backing for why is a bit opaque. We might attribute the social conditions Ulbaek refers to as the reason why language evolved from primate cognition, but more specifics on this matter would be desirable.

The cognitive causes for language, as seen in Figure 1.2 below, are tightly bound to one another. While each factor in the figure has been proposed as the leading (or sole) cause for language emergence, they do exhibit a dynamic connection among one another. For instance, given certain social conditions (e.g. innovation, group hunting, etc), animal cognition could give

way to prospective cognition. Similarly, with the ability to share inner worlds developing (e.g. theory of mind), certain cognitive (planning) and symbolic abilities will affect one another.
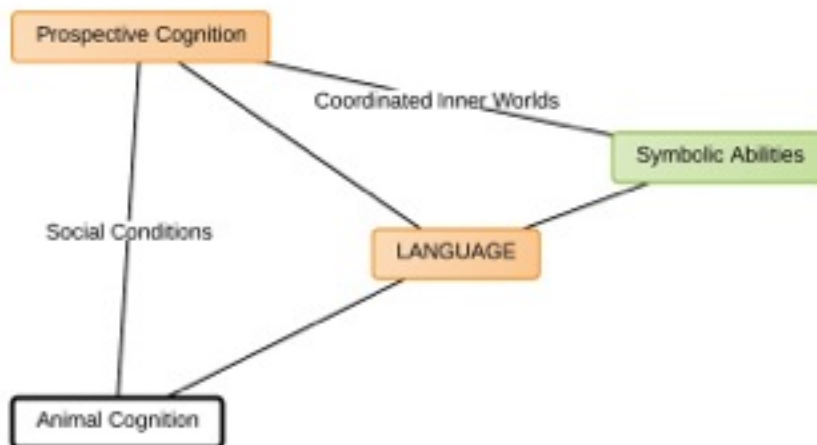
Figure 1.2: Cognitive Causes for Language

Later we will examine in the detail the complex interplay between cognitive factors, but it should be pointed out now that the links between the cognitive factors are in fact socially-based.

BIOLOGICALLY-DRIVEN LANGUAGE EVOLUTION

The final category for these theories of language evolution and emergence are those invoking natural selection as a primary cause, which requires that language be a genetic trait that was selected for in our early ancestors. Language would be considered an adaptation that provided some fitness benefit to its users, and thus was perpetuated in the species.

Pinker and Bloom (1990) claim just that, and are often credited with reigniting the issue of language evolution with their much-debated article on natural selection and language. They claim the human language faculty is a biological adaptation that gradually emerged through natural selection. As a Darwinian adaptation, the language faculty had to increase fitness but

must also have been evolutionarily stable (i.e. reliable). Another tenet of their theory is that natural selection is the only explanation for adaptive complexity- that is, language is complex adaptive design only plausible through natural selection.

Bickerton (2000) (also see Calvin & Bickerton, 2000), modeling his theory from Baldwin effects (i.e. behavioral change bringing about biological changes), proposes a genetically-based cognitive leap that took place in our hominid lineage that allowed for language to evolve in three stages. Protolanguage, he suggests, was the phylogenic precursor to true language, which likely began with *Homo erectus*. Bickerton illustrates protolanguage as simple words and gestures that arose from extractive foraging that was present nearly two million years ago. The cognitive exaptation of reciprocal altruism, a means of performing social calculus (or, "who did what to whom and why," otherwise labeled Agent, Theme, and Goal), is claimed to have been the beginning of syntax. Baldwinian effects then made early syntax more parsable and easier to understand, resulting in a true language that arose within the last two hundred thousand years. Thus, individuals who were better at parsing and processing sentences (the behavioral change) would have been favored by selection (i.e. had more offspring), subsequently improved processing (the biological change). To this end, Bickerton speculates that modern humans may have edged out Neanderthals by having a more advanced phonology (for parsing and creating more words) and perhaps greater numbers for developing culture more rapidly. Therefore, modern humans grew increasingly competent in language, developing "grammatical morphology and parsing algorithms that were incorporated into the human genome by Baldwinian evolution" (p. 282).

Bickerton (2009) has more recently argued for niche construction as a point of language emergence, and not a prior communication system (i.e. non-human animal communication). Niche construction theory (NCT) posits that a species takes part in its own evolution, and for humans the niche that helped form language was culture. Bickerton outlines the process by which NCT explains the rise of language. As early humans engaged in megafauna scavenging the necessity for tracking (tool use included), identification (of fauna or hunting participants), and persuasion (of others to join the scavenging) arose. That is, signal formation would have been a medium through which to accomplish these tasks. To move from megafauna scavenging to power scavenging, recruitment of larger foraging groups had to take place by means of indexical and iconic recruitment signals. Bickerton argues that language (words) and thought co-evolved, while concepts came later. Once humans had language, they were then able to instruct others, engage in social competition and sexual display, gossip, make artifacts, and perform rituals (many of these, as mentioned above, have been suggested as the impetus for language to evolve). These practices would then lead to more language wherein new words would be created to solve practical problems. Similar to how adaptations solve adaptive problems, words solve communication problems.

Similar to Calvin and Bickerton, Wilkins and Wakefield (1995) argue for a reappropriationist model in which structures are modified through gradual natural selection. Reappropriation occurs when a structure/function has reached an evolutionary state that is compatible with and facilitates a new function, and then is modified by natural selection. In the case of language, selective pressure comes from peers after the first generation of language-capable humans. The structure Wilkins and Wakefield propose was modified is the brain,

specifically Broca's area (Brodman 44) and the POT (the intersection of the parietal, occipital, and temporal lobes), which has homologies with and evolved from primates. They suggest Broca's area and the POT were originally designed for motor functions only, and then were reappropriated for processing input into conceptual structures- a pre-requisite for language. In this scenario, language as a communicative function is an adaptation from the reappropriated POT and Broca's Area. This theory falls in line with the type of exaptation categorized by Gould in which exaptations are classified as preadaptions or functional shifts (as compared to spandrels/ by-products)- though this connection is never made explicit by the authors. In contrast with Christiansen and Chater's view of the brain (see above), Wilkins and Wakfield do not suggest that language was molded to fit the reappropriated areas of the brain. Rather, their argument holds that a reappropriated structure (i.e. Broca's area and the POT) will take on a new function for each it has evolved to support, not necessarily that the new function will evolve to fit the pre-existing structure.

In a similar vein, Lieberman (1991) defends a pre-adaptive model of language evolution. A pre-adaption is defined as when an organ originally constructed for one purpose is converted into one for a wholly different purpose. Though, this is distinct from an exaptation in his view. Lieberman argues that a pre-adaptive link between manual motor control and speech was forged by mechanisms in the brain. The speech production mechanisms in turn were a pre-adaptation for syntactic processing mechanisms. Natural selection, then, developed brain mechanisms dedicated to syntax. Setting aside whether or not a dedicated mechanism for syntax exists, a critical issue with the pre-adaptive model is that it does not consider that all pre-adaptations must have first been adaptations, and might then be labeled exaptations.

Botha (2003) problematizes accounts of language evolution through exaptation (including Bickerton, Lieberman, and Wilkes and Wakefield), arguing that these theories lack congruence with Gould and Vbra's (1982) standard of exaptation. Not only do the theorists in question not mention Gould and Vbra's model as a framework, they also seem to confound the terms and theory of exaptation. These theories of language by natural selection fail to account for the richness of human evolution in terms of social and cognitive factors, and attribute our language advancements to fortuitous propagation through selection processes.

Finally, with the advent of modern technologies that have granted researchers access to new data, Stromswold (2010) provides a genetic account for the evolution of language. She explains that if we are to consider genes in the evolution of language, we must look at the heritability of language components (i.e. the phenotypic variation in language, or linguistic ability). If language ability is variable, Stromswold proposes better linguistically "fit" individuals may have reproduced sooner, thus having produced more offspring (and gene copies) throughout their lifetime. Twin studies on linguistic variance reveal language components might only be heritable in a broad sense. That is, there may be influences for gene dominance or interactions between genes, but that the heritability is likely not specific to language. Stromswold drew some implications from findings in genetics and language including that genetic overlap of linguistic and oral-motor abilities could support theories of language evolution from non-linguistic oral abilities (e.g. Lieberman), fine motor abilities (e.g. Bickerton), and social abilities (e.g. Worden, Dunbar). Evidence still remains inconclusive about the extent to which genes and language are directly related, though as technologies advance new knowledge may shed light on the subject.

Figure 1.3 below depicts the interplay between the biological factors outlined here. We can see that while natural selection is the essential head of the biologically-driven theories of language evolution, the proposed paths to it are not necessary the same. Pre-adaptations as well as reappropriated structures might not originally have been under selective pressure (for language, at least), but eventually come to be so. In the next section we will see how other causes will contribute to these biological determinants.
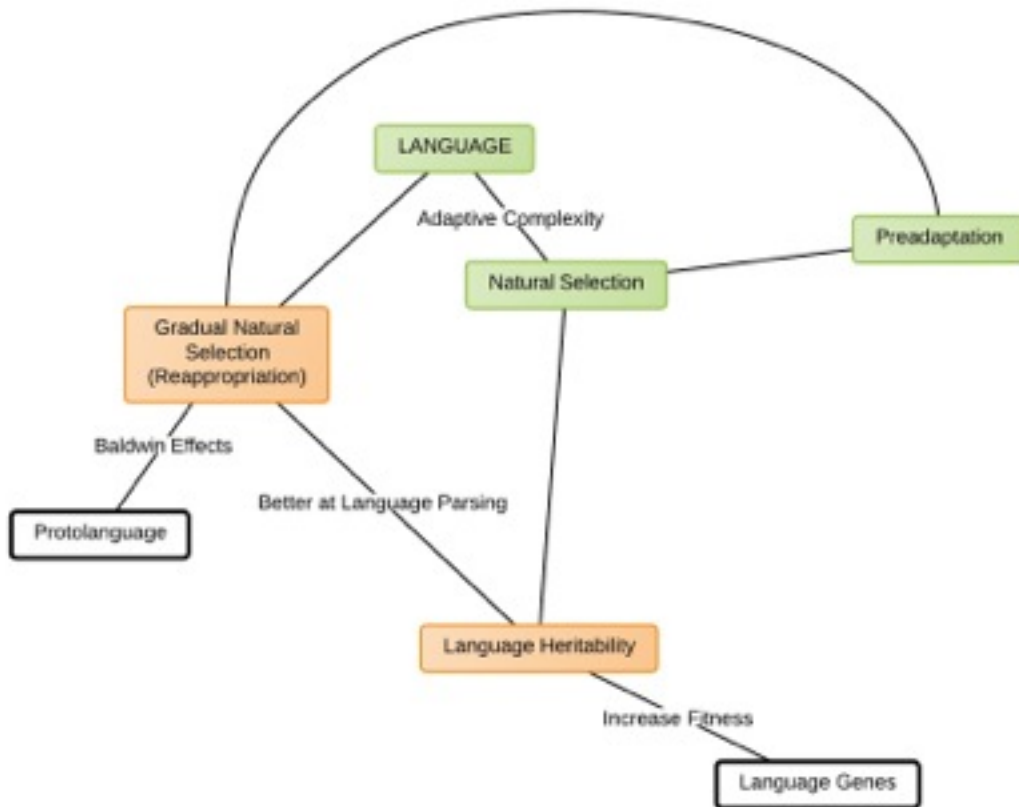
Figure 1.3: Biological Causes for Language

In sum, the three perspectives on language evolution and emergence highlight the differences in the how, why, and when questions of these phenomena. Even within a certain

frame, theories tend to contradict one another (given, a few do concede that other theories are compatible with one they propose).

LANGUAGE AS A COMPLEX ADAPTIVE SYSTEM: A THEORETICAL FRAMEWORK

Language as we know and use it is an incredible accomplishment and feature of human society. Its emergence and evolution is therefore of great interest to linguistics and researchers in seemingly distant fields. As mentioned earlier, the goal of this paper is not to discount others' theories, but rather propose we work from a framework that might suggest language is complex and its emergence would rely on the chaotic interactions of diverse factors. Lee et al (2009) outline the chaos from which patterns emerge through self-organization. First there is an equilibrium state in which no outside, environmental factors influence the state. In the near-equilibrium state experiences weak external influences (e.g. from the environment), though patterns have yet to emerge. Finally, in the non-equilibrium state, there is active chaos and strong environmental influences control the system and patterns emerge. We might consider language as the pattern that emerged given the increased chaos of external factors, including social, cognitive, biological, and neurological conditions.

Lee and his colleagues believe language is a complex adaptive system (CAS). Following the three principles below, I will discuss these characteristics in terms of language emergence and evolution theory.

1. "CASs are systems of complex structures in which patterns emerge dynamically through local interactions among many agents in spite of the absence of pre-ordained design" (pp. 17-18).

Lee et al propose language as a CAS in terms of the emergence of linguistic structures, wherein the "agents" are not only the human interlocutors present in the late Pleistocene, but the number of words/representations present to create linguistic structures such as grammar and phonology. I do not find fault with this interpretation, but I do believe this same first principle might also have relevance to language emergence in a more general sense (i.e. communicative origin of language). Consider that the "many agents" above refer to the pre-requisites many theorists have proposed in language emergences theories. These agents might include certain cognitive abilities like theory of mind, social practices such as grooming and information sharing, neurological changes, anatomical changes, and possibly other biological factors. The dynamically increasing repertoire of factors influencing early humans would no doubt be acting on one another in a complex interaction of abilities and experiences.

Furthermore, the "absence of pre-ordained design" falls in line with many theories about language not being selected for in a complete and fully functional way (c.f. Chomsky's camp on a innate, biologically-based Universal Grammar that sprung into being in complete form). Language, many believe, has been evolving since its first generation of users due to its tightly bound relationship to culture. In a general sense, language may not have a "pre-ordained design," rather language has been adapted to fit our human brain, our anatomical structures, and our functional uses of it. That is, genes may not have necessarily pre-determined language at any point in human evolutionary history. Rather than be instantiated by genetic code, the agents posited above would interact to generate language and could keep doing so throughout human history.

Furthermore, in later chapters here, we shall see how systematic languages emerge without a predetermined communication mode given to its intended users (see Cornish, Tamariz, and Kirby, 2009, for a more detailed account).

The first principle then is a perspective on the emergence, and even possibly the subsequent evolution, of language. In this scenario, multiple agents would contribute to the emergence of language, and the lack of one may have resulted in a very different emergent pattern. From this viewpoint, language could have emerged from multiple causes interacting in chaotic concert with one another.

Multistrata, the hierarchical building of agents, is proposed as part of the first tenet as well (Holland, 1975, 1995, as cited in Lee et al, 2009). The hierarchy of agents relies first on the most basic agent as an initial building block. Subsequent agents then increase in complexity at each new level. Current language evolution theories already follow this model, though possibly unwittingly. Bickerton, for example, draws a timeline for the increased complexity of language over two million years, beginning with protolanguage until the genetic blocks, in his scenario, are built up to true language. Lieberman uses a similar model for the advancement of the speech system, and Corballis does the same for the transition from gestural to vocal language. Considering the building blocks, though, does not require us to regard one pathway as the only possible option for language emergence. In fact, the multistrata model permits the development of an elaborate evolutionary history (e.g. giving the complete story).

2. "Small inputs into a CAS can cause major changes" (p. 18).

Inputs, in an evolutionary context, can be translated in many ways, and probably should be in order to account for the variety of inputs that could have contributed to language. Inputs into an evolving CAS, such as language, can include social factors (increased group size, competition and coordination, information sharing), cultural evolution (tool innovation, art, music), physiological changes (lowered larynx, brain size, bipedalism), and possibly genetic changes (re-organization of the brain, genetic drift, gene mutation). Regardless of the camp one belongs to with respect to these inputs and their relationship to language evolution, it is important to note these inputs to language emergence ("the major change"). To provide an example, the input of a lowered larynx changed human anatomy so that a wider range of sounds was then producible. Without this small input, language would not have formed as it did.

Non-linearity is a central aspect of this second principle. In linear interactions, small inputs can only generate small changes. Non-linear interactions, in contrast, can demonstrate large effects from small inputs or variables. Thus, we are to consider the product of the variables, rather than their sum. We could approach this by hypothesizing about how certain variables may interact with one another and what their collective interaction brings about in the CAS, rather than reasoning one variable plus another led to a particular change. That is, we should create a theory of language evolution in terms of multiple causalities, wherein various inputs (variables) interact in a non-linear way to produce an outcome, in this case, language.

27

3. CASs "show a general tendency for 'coherence under change'" (p. 18).

Even with a great number of inputs that can cause changes to the system, language is still language. Take, for example, Corballis' account of gestural language evolving into vocal language. Is there not still coherence? Is the communicative intention not the same? Another example of "coherence under change" appears after the emergence of language when language has to become understandable, producible, and learnable. Through the pruning of linguistic features that are not suitable for the human tongue, brain, or society, we arrive at a more coherent communication system that has been adapted to our abilities to use language and our functional uses of language. In this sense, we may not view language as a biologically endowed ability, rather a cultural system that has been adapted to human brains (understandable and producible) and transmitted culturally (learnable).

Language as a complex adaptive system entails multiple variables being built upon and interacting with one another, contributing to language emergence and evolution (in stark contrast to Pinker and Bloom's argument that complexity only arises from natural selection). This approach to theorizing about language origins seems to be a viable option for proposing not only a complete evolutionary history, but necessitates thinking about language as a complex phenomenon requiring multiple causes for its origin.

Below Figure 1.4 represents the causes proposed for language (refer to the literature reviewed above), the "roots" of such causes, and the connections between them. Here we can glimpse the complexity of diverse interactants coming together to build up a single phenomenon: language. Social, cognitive, biological, cultural, and neurobiological factors could all have a role

in the emergence of language. Furthermore, each one is intricately linked to another. Social

abilities could be transformed into biological traits. Cognitive capacities could support (and be

supported by) social behaviors. Culture could not only benefit sociality, but cognition as well.

The links to be made between the multitude of factors leading to language are innumerable, and

could grow with more advances in research.
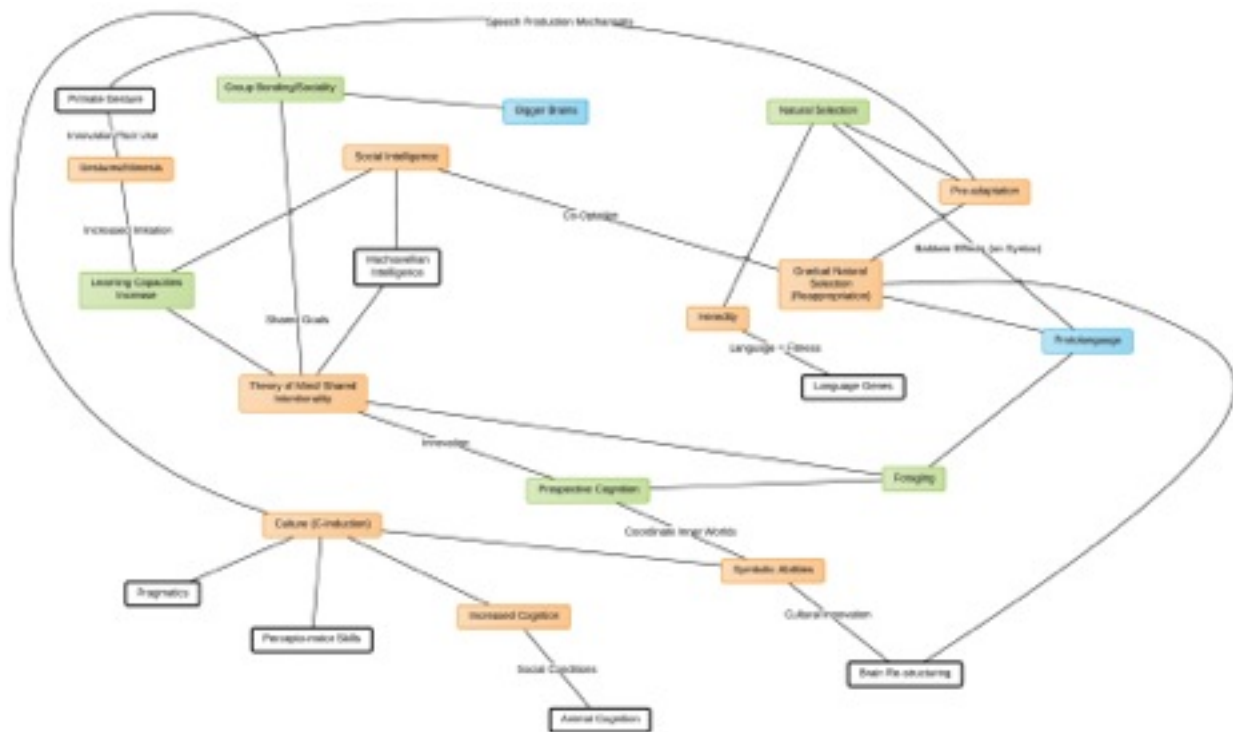
Figure 1.4: The Relation of Causalities for Language Evolution

It would be difficult to parse any one of these "causes" for language, as many rely on and support

one another. The interaction among these seemingly independent agents is the chaos that brought

about language.

The Value in Considering Multiple/Non-linear Causality

Through this review of theories on language emergence and evolution, it becomes clear that researchers are often using quite a limited scope to explore this complicated topic. Here I have proposed that future hypotheses and theories on language origins consider the CAS perspective for language origins. If language emerged from multiple "causes" and complex interactions requiring social, cognitive, and physiological features to be working in concert to produce a viable communication system to coordinate social action and everyday life in the EEA, considering a framework such as the complex adaptive system would allow researchers to not only consider the non-linearity of evolution, but also provide an account of the evolutionary how's and why's.

Unfortunately, many researchers limit their claims of language origins to one (sometimes a slight combination) of factors including oral motor skill, fine motor skill, non-verbal cognitive abilities, social cognition, and theory mind/mind-reading. But it is entirely plausible that multiple non-verbal abilities were co-opted for or underlie language. If we consider language as emergent from a chaos of converging aspects of human evolution, we could then concede that there is no single cause for language, and probably no single set of causes.

To conclude, anthropologist Daniel Everett (2012) has recently published an account of three "platforms" that converged to spawn language among our early ancestors. These platforms include the cognitive (intentionality, directedness, theory of mind, consciousness, and culture), the physical (the human vocal apparatus), and the cerebral (physiology and structure of the brain). While Everett does not elaborate on how these platforms might have interacted, he has

definitely provided a "starting off" point for considering a multiple-causation scenario for

language origins that derives from diverse yet related aspects of human evolution.

Notes

[1] Barrett et al (reviewer comments in Chater and Christiansen, 2008), while agreeing with the aspect of cultural evolution Chater and Christiansen propose, argue that natural selection can, and does, in fact act on moving targets in which statistical regularities are present. They further argue for the co-evolution of genes and culture, with respect to language.

[2] The directional arrows in these figures are not meant to indicate a temporal scale, nor is it the case that the arrows reflect a continuous flow from one factor to another. The arrows are simply meant to illustrate how one factor may contribute to, support, or lead to another one.

CHAPTER 2

THE NATURE OF LANGUAGE IN INTERACTION

When two individuals set out to complete a task or goal, communication greatly facilitates the accomplishment. There might be numerous reasons for people to work together to achieve goals or attain certain outcomes, from task difficulty to the need for multiple hands for manipulation, from novice inability to camaraderie. Engaging in action with others necessitates cooperation, best carried out by humans through various communication systems (vocal, gestural, non-linguistic) that encompass and rely upon other discourse features to bring about meaning. Co-operative action refers to the joint construction of action by interacting individuals performing operations on diverse materials (Goodwin, 2013). The organization and co-operation of human action allows people to coordinate activity in ways that might not be present in other animals. People engage in joint action by bringing together meaningful materials (e.g. language, cultural artifacts), and performing operations on a prior, public substrate (e.g. action, talk). Moreover, the organization of human interaction is simultaneous and embedded in context. Naturalistic data best demonstrate these features of human co-operative action.

In the data to be analyzed below, a grandson and his grandmother interact in everyday activities to accomplish a number of actions, including counting objects and prepping meals. These Spanish-English bilinguals demonstrate how communication is aided by the integration of the environment, body, and other semiotic resources. A close examination of this data will reveal how talk works with multiple resources to bring about and carry out co-operative actions (Goodwin, 2013), which speaks to how we use language and other discourse features to effectively communicate.

32

METHODOLOGY

Naturalistic data best demonstrate the nature of language in interaction. It provides an observational glimpse into how language is constructed and complemented with other features of the environment. A qualitative investigation of language use is necessary for a comprehensive analysis of the nature of language.

For this study, the data collected was on human participants. Specifically, focal follows were conducted with a Spanish-English bilingual family. The family consisted of two grandparents and their grandson. The grandparents were approximately sixty years of age, and their grandson 28 months. The grandmother is a native Spanish-English bilingual, as is her grandson. The grandfather is native speaker of English, who demonstrates high fluency in Spanish as well. Their daily routines were video documented for a number of days and relevant data were transcribed (following the Jefferson model, see Appendix A for transcription guide). Examining the nature of interactions and how language aids interactions may give us insights into the reasons for such a complex communicative system to develop in humans.

DATA EXAMPLE #1

Here we examined the dyadic interaction between a Spanish-English bilingual grandmother and her grandson. The specific interactions analyzed involve counting discrete objects during "reading time." The data set includes counting actions with two different books. In the first encounter with counting discrete picture-objects, the grandson does not produce an audible, nor accurate, answer. His counting points are imprecisely performed, and he does not seek the assistance of his grandmother. In a subsequent attempt, initiated by the grandmother, the grandson still does not perform the count accurate (he "over-counts" the objects). With the aide

of his grandmother, in a second attempt, they both jointly construct the action to reveal the accurate count (transcripts for the talk in each set are provided in the Appendix A). As previously discussed, human action is built co-operatively and complexly, and as such, we investigate a multitude of features that contribute to human interactional behavior. These features include the positioning and organization of the body, the simultaneity of actions, use of semiotic resources (e.g. cultural artifacts, tools, and language), environmentally coupled gestures, and talk.

Organization of Action

The organization of the participants, namely the position of their bodies, creates an embodied framework for the coordination of action. Lap sitting is a common child-caregiver position, both in human and ape groups. It allows the child to feel the movement of his caregiver, and also to attend to objects outside of the dyad. In hunter-gather cultures, lap sitting is commonly employed so as to expose the child to many inputs and feedback (not just from the caregiver, usually the mother). The lap sitting position here allows for many aspects of coordinated action to occur. First, it provides a visual frame in which the novice and expert can attend to the same object; that is, they can have the book (on the floor in front of them) as a common focus (as seen in Image 2.1). With their gaze situated in the same direction, the object to which they attend is obvious and made relevant. This common focus is similar to Tomasello's triadic joint attention wherein two participants will gaze upon an object- the object of their joint attention. From the join attention, the participants can operate on the object and understand that they are both attending to it.

Image 2.1: Common Focus in Lap Sitting   **Image 2.2: Manipulation of novice's body by expert**

Second, the intertwined nature of lap sitting allows for ease of manipulation of the novice by the expert. When the novice displays incomplete knowledge of how to perform the action, the expert is able to manipulate and adjust the action, in a way modeling the "correct" method. In this example, the action of counting discrete objects is incomplete (as will be seen below) in that the novice's imprecise points and matched counting are disjointed. The expert, the grandmother, is able to hold and manipulate the grandson's hand from her position (Image 2.2). The manipulation allows for a mapping of the count number to each discrete object on the page. The touch, made possible from the lap sitting position, enables the simultaneous performance of action as well (to be discussed below).

Third, in support of the pedagogical nature of this interaction, the lap sitting position allows both the "teacher" and the "learner" to establish a common referent. Not only does the positioning disambiguate the focus of the pedagogical interaction, but it also makes what is to be learned less opaque. Although the grandson cannot readily see/know where his grandmother is gazing (this also becomes more apparent with pointing, to be discussed below), he does have

access to her body movement- through the closely intertwined nature of lap sitting- and where

she might be aligning herself to in the environment. Also, given the labeling of this activity as

"reading time," it might be recognized as an appropriate setting for which knowledge might be

passed from one person to the other. That is, situated activities (Goodwin and Goodwin, 1996),

such as "reading time," are spaces in which cognitive tasks can be carried out, including these

pedagogical tasks (Csibra and Gergely, 2006). Thus, the learner might be more inclined to

interpret the actions occurring here as pedagogical, and requisite of a common referent upon

which to act.

Co-operative Action

It has been posited that the co-operative nature of human action might be distinct to our

own species (Goodwin, 2013). As such, humans are able to carry out and organize action through

unique means, including engaging in simultaneous action. Performing actions simultaneously

allows for multiple parties to operate on an object. For instance, in the data example in Image

2.3, by holding the hand of her grandson, the grandmother performs the counting action with her

novice co-participant. Simultaneous action is particularly important in learning (in this case,

counting); with the guidance of the expert teacher, engaged in the same movement and the same

time, the novice is able to detect the precise movement of the hand (specifically, the deictic

point) to each discrete object. If, for example, the grandmother had performed this same action

sequentially, that is, done the discrete objet points herself as a demonstration and then had her

grandson repeat the counting, he would not benefit from the embodied learning that would have

taken place with the simultaneous pointing. Simultaneity, then, exhibits a greater degree of

collaboration in the social action, and may also induce an affective response. Such responses may be responsible for learning and memory (Schumann, 1997).

Image 2.3: Simultaneous Counting Sequence

Furthermore, the simultaneous action between the novice and expert leads to the construction of a competent member of that society. As a novice who has an incomplete version of the action (liminal state of knowledge of action), the expert must guide the novice in learning the appropriateness of actions and how to perform them. The pedagogical implementation of simultaneous actions does not supply an example alone, but rather provides an embodied demonstration, enacted by both the expert and novice. Examples of this type of social learning are also present in hunter-gather societies (Hewlett, 2013), suggesting its ubiquity in humans. The simultaneous pointing of the grandmother and grandson allows the grandmother to demonstrate the preciseness that deictic points in counting should employ.

Use of Semiotic Resources

Humans build action by bringing together and making use of multiple resources and materials. Since human action is situated in a given environment, and builds from materials within that environment, actions thus become meaningful when considered in conjunction with their surroundings. Here we see the action relies on and is sedimented in the book, its pictures, and the discrete objects in the images. These semiotic resources make the action of counting

relevant, non-opaque, and meaningful. Furthermore, the book and its images provide a substrate (to which both participants have access to) that can be operated upon. That is, the grandmother could ask about the color of objects, the characters in the story images, and the number of a certain type of object. The book is an open tool (Hutchins, 1993) upon which operations can be made directly on its surface and are visible to both participants.

Environmentally coupled gestures (Goodwin, 2007), then, allow for the particular objects in the setting to become relevant with deictic points. The grandmother first indicates the specific object that she would like her grandson to count with a deictic point to the babies in the book image. She couples this gesture not only with the book image, but with the question "Quien se va a subir al tren?" ("Who is getting on the train?")(line 1, Image 2.4). The combination of these resources gives the grandson a platform for action. With the environmentally coupled point, he has access to the object in question; and, the question requests an action of him- specifically, to identify the object that has been indexed with a prior deictic point. Thus, the environmentally coupled point makes not only an object relevant for the subsequent action, but also the exact action to be produced.


```
01  GMA:    Quien se va a subir al tren,
02  GSN:    Baey-bie:[:s
```

Image 2.4: Deictic point coupled with talk and indexing the referent within the semiotic resource

38

Another semiotic resource that contributes to this interaction is the prosody that inflects the talk. For instance, the grandmother leads the pedagogical interaction, but encourages her grandson to join in the counting effort. When he is initially unable to produce the exact count, the grandmother holds the grandson's hand in the point (to be analyzed in more depth below) and simultaneously begins the count with an elongated and rising "U::no::" (see Image 2.6 below). The prosody in the grandmother's talk is a cue to the grandson to pick up the talk as well; that is, collaborate in the simultaneous counting effort.

Language and Action

The features discussed above work in concert with one another to build actions that encourage human sociality and learning. Human language is a unique feature to the organization of human action, and as such allows human co-participants to coordinate action in ways distinct from nonhuman animals. While the distinctions between human and nonhuman interaction will be discussed below, the use of language will be further elaborated here. Language couples with resources in the current, even past and future, environment to bring about relevant action to complete shared goals of individuals in a community.

From a conversation analytic perspective, the actions performed in these sequences are question-answer adjacency pairs. While a deictic point (Image 2.5, below) may help learner attend to the appropriate object/referent, the talk (i.e. language) makes the desired action known. The answer can be provided through various sets of practices. For example, the grandson could provide a verbal, or even gestural, answer without a counting sequence, or he could do as he has done here and initiate a count action. Moreover, the count action is best achieved aloud, using number units, as opposed to silent, or mouthed, counting. Counting aloud also allows for better

monitoring, as the learner's mental state is then displayed publicly for others to build upon. Furthermore, the question-action is a means for the grandmother to teach her grandson to become a competent member of a community of pre-school-age children. Her questioning jointly delivered with a deictic point makes relevant the next action: a response. The response her grandson provides will allow the grandmother to monitor the novice's progress towards becoming a knowledgeable participant in counting activities. When the grandmother notices the liminal state of the grandson's knowledge of counting- not providing a one-to-one mapping of the number to the image- she is able to intervene and coordinate the action simultaneously, as though she is providing a model for future use.

Pedagogy and Social Learning

Csibra and Gergely (2006) posit that humans are adapted to engage in pedagogical interactions to receive knowledge from and transfer it to conspecifics. As a type of social learning, pedagogy requires three crucial design features to permit knowledge transmission. First, the expert (or teacher) must mark the action as a pedagogical one, that is, one that is meant for a particular pupil to learn something through demonstration of expert knowledge. Language is helpful means (among humans) to indicate that an action is meant to be a "teaching moment;" however, other ostensive stimuli such as eye gaze and touch can also communicate the pedagogical intention. For example, in the Image 2.6 below the grandmother does not use language to indicate the pedagogical nature of the action, but rather holds the grandson's hand in the deictic point to model the appropriate behavior.

Second, the expert must specify the referent of the pedagogical interaction, so that the novice attends to the relevant resources in the environment (Csibra & Gergely, 2006). While this

could be established with deictic gestures, body position could also facilitate the identification of a specific referent. The grandmother and grandson's body positions (Images 2.1 and 2.2 above) allow them both to attend to the same object in their visual field, the book. Furthermore, the labeling of this time as 'reading time' focuses both the expert and novice in a situated activity, from which they can perform other relevant operations. Third, the expert should disambiguate what is relevant or not to the transfer of knowledge (Csibra & Gergely, 2006). This requires an assessment of what the learner lacks, which can be achieved through monitoring the completeness of the knowledge. Disambiguation can be accomplished through talk coupled with gestures and other semiotic resources. In the grandmother-grandson dyad, the first step to disambiguation is the expert giving a deictic point, coupled with the talk "Quien se va a subir al tren," ("Who is getting on the train" transcript 1, line 1- Image 2.4) to indicate the discrete object on which to operate. Her next question, "Puedes contar los babies" ("Can you count the babies" transcript 1, line 3), requests the counting action; however, the grandson only provides indiscriminate counting (under his breath, and mouthed, not spoken) and imprecise pointing (Image 2.5). The counting action here never comes to fruition, but in a later instance of a request to count, the grandmother notices the incomplete ability of her grandson to point to the relevant objects with a precise number, and intervenes to disambiguate further. The second step, then, involves the grandmother jointly pointing with the grandson to discrete objects while simultaneously starting the count sequence (Image 2.6). It appears that the design features necessary for pedagogy are present in this data, however, there may be more that constitutes social learning in human interactions.

```
03  GMA:    Puedes contar los babies,
04  GSN:     Un-° ((mouths numbers 2 & 3, indiscriminate
                 pointing to count objects))

15  GSN:     w^w:on::e ((pointing to first count object)) (.)
                 two°
16  GSN:     ((moves point with each count object))
                 ((inaudible mouthing of numbers))
                 five° six° seven°
```

Image 2.5: Indiscriminate pointing leads to imprecise counting

uno^ os tres cuatro cinco seis oto   uno::: [os   tres cuatro] cin-

Image 2.6: Imprecise counting/pointing to guided, joint counting/pointing

Hewlett (2013) also investigates the ubiquity of social learning and pedagogy among human cultures, namely in the Aka forager society. There he found that demonstration, among many other pedagogical resources, was present among hunter-gatherer teaching/learning. Aka mothers and fathers would readily demonstrate how to perform socially accepted practices by simultaneously enacting them. For example, while a child manipulates a knife or machete, a parent will jointly hold the child's hand and the tool to guide the appropriate action. This way the child can understand the proper way to hold a certain tool, the force to exert onto it, and the acceptable way to use it. The jointly coordinated action arises from the monitoring the parent does while the child experiments with the tool. Furthermore, the children of the Aka foragers are often sat in the laps of their caregivers, facing outward toward the group, in order to receive the

42

more input on appropriate behavior and action. Similar to the interaction above, lap sitting allows for children to attend to objects their caregivers manipulate, as well as be accessible to caregiver manipulation. The widespread nature of this form of pedagogy implies that humans have adapted to organize their action in such a way that builds competent members of a community and to transmit knowledge (e.g. culture) vertically and horizontally.

Tennie, Call, and Tomasello (2009) argue that learning in humans is distinct from other primates in that human infants/children focus on the process, not the product of pedagogical interactions. Their argument is evidenced by the imitation that is done of demonstrators by human children, compared with chimpanzees, specifically the imitation of non-essential actions to reach the desired outcome. If we consider the pedagogical interaction of the data presented here, it is apparent that the focus is not on the product (the number of objects, such as honey-pots or babies), but rather the process of counting (the practices involved in counting/learning to count). Tennie et al note that children are careful to attend to "the way [the task] *should* be done," and that they are quick to chastise those who do not comply. Not only are children attending to the process (whether the actions be necessary or not), but they are also keen on imitating, or "being like," their conspecific demonstrator. The affiliative drive in human infants has been posited as a means for language acquisition (Lee et al, 2009); however, it may be generalizable to acquisition of new knowledge in many domains. The high level of human cooperation, driven by a desire for affiliation, allows for active teaching in which children learn to conform to the social practices of their community.

DATA EXAMPLE #2

In this second excerpt of data, the grandson is eating a meal at his chair and has enlisted his grandmother to help assemble the food for ease of eating. The meal consists of beans, rice, avocado, and tortillas. It is not clear the precise purpose for inciting his grandmother's participation in his eating practice; however, many factors may be at play. For example, his hands may not be large enough to manipulate the materials (or he may lack the manual dexterity), or he may be seeking a bonding moment through more social behavior. After his grandmother proposes some options for eating the meal (that is, what should be eaten and how), the grandson repairs her efforts to be more aligned with his own expectations. The talk is as follows (translation provided in Appendix A):

> (1) "Pedacitos"
>
> 20　GSN:　Sí nan- ASÍ NO nana
> 21　GMA:　Ah cómo te lo hago oh uh ((GSN holds edge of tortilla))
> 22　GSN:　(Part) ((moves hands together and apart to represent tearing motion))
> 23　GMA:　Así mi- así
> 24　GSN:　NO:: ((holds hands up in 'stop' position))
> 25　GMA:　Cómo lo quieres
> 26　GSN:　Así ((makes tearing motion with fingers))
> 27　GMA:　Oh en pedacitos
> 28　GSN:　Sí::

These series of corrections (lines 20, 24, 26) are exemplary of how language is used in conjunction with other resources to make adjusts on actions others' do, but in the context of goal attainment by the repairer.

Talk and Gesture

To demonstrate how he wants to eat his meal, the grandson employs coordinated talk and hand gestures to convey his desires to his grandmother, who is manipulating the food for him. After being chastised for the incorrect folding of the tortilla (line 20, in example 1a/b below), the

grandmother asks how she should be preparing it (line 21) while the grandson holds the edge of the tortilla (Image 2.7) as a means to demonstrate pulling it apart. He is performing the tearing action on the tortilla initially, and then proceeds to enact the tearing by pulling his hand together and apart (Image 2.8).

Image **2.7: Tearing Enacted on Object**   **Image 2.8: Tearing Enacted away from Object**

Grandmother's second attempt is also met with repair (line 24) with the exclamatory "No" and a gesture wherein the grandson holds up his hands, indicating the current action be terminated. Following yet another request for explanation from his grandmother (line 25), the grandson again enacts the tearing motion with a simultaneous "Asi/Like this" (line 26), displaying how the action should be carried out to achieve the goal, in contrast to enacting the actual goal ("pieces of tortilla," Image 2.9). In line 27, the grandmother discovers the desired eating style of her grandson, making it clear by not only putting a label to the goal ("pedacitos/ little pieces") but moving forward with the new action of tearing the tortilla into pieces.

Image 2.9: Re-enacted Tear Motion

Here, the talk is enhanced with gesture, which becomes particularly useful when the label

(lexical item) is not readily available to one, or both, of the participants.

Talk and Eye Gaze

Actions are not accomplished by talk in isolation. Monitoring of the understanding of

vocal output is essential to progress actions. In this sequence, we have seen how the grandson

simultaneously uses talk and gestures to describe a desired action that his grandmother should

perform. He also monitors her comprehension of his repairs, particularly at the second repair, by

turning his eye gaze to her as he completes the gesture (Image 2.10). Initially, his gaze is toward

the object being manipulated (and the action being done by the grandmother) to bring attention

to this point of repair. However, as he completes his repair (with talk, "Así (like this)" and a

pulling gesture), he shifts his gaze to look at this grandmother's face, seeking alignment (Stivers

& Rossano, 2010) to the desired next action.
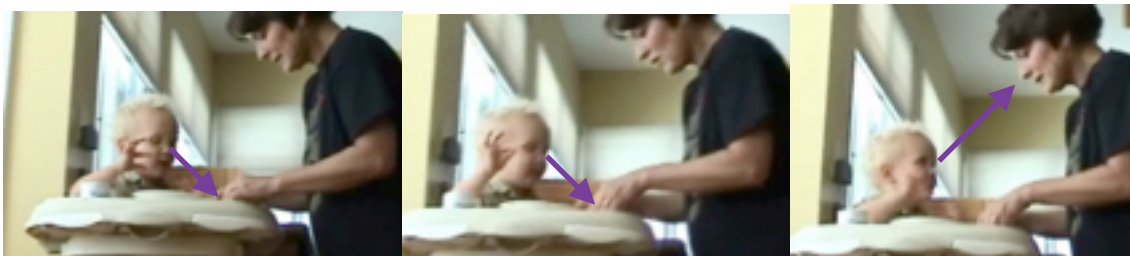
Image 2.10: Raising Gaze near Gesture/Talk Completion

To be sure his grandmother has read his intentions accurately, the grandson raises his gaze to

look at her face for affirmation, which she provides with an "Oh"- a token of recognition.

Furthermore, the acknowledgement of the desire goal, which her grandson has been attempting

to communicate through the combined use of talk and gesture, is given with her offer of the label

"pedacitos (little pieces)" (provisioning of the word in a word search, Goodwin & Goodwin,

1986) and movement to produce the requested outcome. Here we see how a novice participant

can use limited language in combination with gestures and eye gaze to achieve greater communicative success than with just language alone.

POSSIBILITIES FOR LANGUAGE EVOLUTION

To carry out successful communication, we combine talk with gestures, meaningful eye gaze, and the use of objects. These resources can disambiguate the talk (and corresponding action) and help to progress the interaction. While the actual language might hold much expressivity, incorporating other aspects, such as gesture, can enhance it. Considering the evolution of language into a complex form of communication that is highly symbolic, it is suggested the integration of other features into our communicative interactions could allow for a more complex and symbolic communicative system to evolve, as problems of communicative success could be mediated by such integration. I also propose the ability to combine talk and interactional features is a learned phenomenon, wherein novices become competent participants in interaction through practice and observation.

Young users of language are not, however, excluded from combining resources to accomplish their needs and goals. As the above data exemplifies, young children make use of and employ varied resources to interact with others. The ability of children to incorporate talk, gesture, and the environment- given their possible disadvantage with lexical terms- is remarkable and speaks to the centrality of such an ability to human interaction. Furthermore, this data demonstrates how talk is not used, nor considered by others, as an isolated feature, but rather one that works in concert with other semiotic resources.

The examination of the organization of human action faces a similar problem as the research on language emergence/evolution: elements are usually considered in isolation. In the

47

recent wave of language evolution research, theorists (see Chapter 1 for a review) have attempted to draw one single cause as the impetus for language emergence, from natural selection to social brains to gestural imitation. Unfortunately, this linear thinking has allowed little space for researchers to consider the complexity of human action, society, and cognition that could contribute to the development and evolution of human language. In a similar vein, human action research tends to focus on elements in isolation as well; consider the focus on talk in conversation analysis. Gesture is another feature that is often examined without consideration for the accompanying talk or local context. The building of action and language cannot be constructed from a single cause or feature, but rather requires the complexity of multiple agents and resources acting together.

Complex adaptive systems are those in which phenomena emerge from multiple agents being brought together and engaging with one another (Lee et al, 2009). With respect to language, "agents" are not necessarily animate (e.g. people), but could also be the various causes underlying its emergence, including increased sociality, theory of mind, biological adaptations to the brain and production organs (vocal apparatus and gesture), culture, and abstract thought. If we are to view language, as a communicative resource, through this lens, a close examination of its modern use gives us a glimpse at the nature of the phenomenon. Then, it is plausible to draw parallels between modern use and primitive use of language. The everyday actions we perform with language exhibit a complexity that is quite distinct. We bring together talk, gestures, eye gaze, tools and objects, and the environment in which the activity is situated to make meaning, engage in interaction, and accomplish goals. These features of language in interaction would

likely benefit our ancestors as well. Therefore, as proposed earlier, language not used in isolation

likely could not have emerged in isolation.

CHAPTER 3

STUDY 1: GESTURAL COMMUNICATION OF MOVEMENT

INTRODUCTION

In the language evolution research, interaction is a feature that is not widely investigated nor considered for its role in emerging and evolving communication systems. Some studies parse aspects of interaction, while other eliminate it completely. This is problematic given the nature of human communication and language use: a highly interactive and co-operatively built activity. Furthermore, cultural transmission, paramount in many language evolution models, is inherently interactive, requiring the learning (either direct or indirect) of systems, patterns, and traditions by naive participants from competent members of a group. The complication with studying interaction, though, is that there is much to consider: people (who they are, what they know, their status), context (the environment, semiotic resources), communication structure (turn-taking, adjacency pairs, repair, overlap), and other discourse features (gestures, prosody, eye gaze, body orientation and movement). We cannot deny the environment in which we use language, who is using it, and what accompanies it. The current study will investigate, experimentally, the evolution of a gestural communication system and how certain interactional features might facilitate this evolution.

To date, there have not been many systematic investigations into the evolution of language that rely on human participant paradigms, as computer modeling has dominated this field of research. Kirby (2002) follows an iterated learning model, a computer simulation of language evolution over generations, which is based in vertical transmission of language. Language, as it goes through generations, is "incrementally influenced by agents' learning biases

until the language reaches an equilibrium" (Fay, Garrod, Roberts, & Swoboda, 2010, p. 352). However, given the emphasis on vertical transmission, agent interaction (horizontally) is not considered in this model of language evolution. In the model, "children" receive language input (a set of meaning-signal pairs generated from a random subset) from "adults," from which they deduce rules. Once the "child" reaches the equivalent of the critical period, he becomes an "adult" who influences the next generation of learners. While based on models of cultural evolution (see Boyd & Richerson, 2005), in this model "children" are not able to influence the language of "adults."

More collaborative models have emerged in the field, most notably from Steels (2003). Unlike the individualistic models involving vertical transmission, collaborative accounts incorporate horizontal (and bidirectional) transmission of language, which emerges from dynamic social interactions of its agents. For example, in the "talking heads experiment" simulated agents (in pairs) must resolve a coordination problem wherein new members must align with an existing linguistic system (Fay et al, 2010). While miscommunication may occur initially, the agents eventually settle on words for referents in a self-organizing process. Although, due to the simulated nature of these experiments, it is difficult to consider the social contexts or dominance hierarchies that might impact the maintenance of particular referents over others. Nonetheless, computer modeling (e.g. simulations) of language evolution have proven useful to the field, and have, in fact, generated participant-based designs.

One such natural experiment (i.e. not computer simulated) was conducted by Fay et al (2010), adapting Garrod et al's (2007) "Pictionary" game method. In this paradigm, paired human participants are given a set of "easily confusable" concepts. One person drew the

concepts, while the other identified the drawing with a corresponding lexical item. Over multiple trials, drawings became less iconic and more abstract/symbolic as drawings became less complex, but identification accuracy increased. Furthermore, when comparing isolated pairs to a community of participants, community signs were more accurately decoded and recognized, suggesting a preference for these signed by new "learners." This method does, however, rely on the visual representation of signs, which might prove problematic when extrapolating to language origins theories.

Iterated learning models involving human participants have, in the past decade, become more prevalent in the language evolution research realm. Kirby, Cornish, and Smith's (2008) seminal study of cultural transmission of language through iterated learning demonstrates how language can evolve over generations without a pre-determined design. Here, participants learn an "alien" language which they use to label objects over a number of rounds. The whole language, as learned by a participant, is 'passed down' to the next naïve participant in a diffusion chain that continues over a number of participants. This gives way to language evolution in vertical transmission, however there is not an opportunity for reverse transmission, nor horizontal transmission (as a dyad might exhibit). They found that over generations the number of individual words decreases and that the transmission error decreased as well. This indicates the language became more learnable over generations, and that it had evolved linguistically. These findings demonstrate that language is an adaptive system that conforms to communication pressures and is culturally accumulated.

Researchers have also taken up gesture as a means to study language evolution, as it is also a means of communication. Many studies (Goldin-Meadow et al (2008), Schouwstra (2012))

have investigated the word order in gestures of non-signing individuals, finding that regardless of native language word order, there is a great trend to use the same word (SOV) in gestures. Schouwstra (2012) furthered this research by demonstrating that the verb type has an impact on the word order in gestured phrases, changing from SOV for motion events to SVO for intension events. More in line with the current study, Fay, Arbib, and Garrod (2013) measured the communicative success in three experimental conditions: verbal, gestural, and mixed. They found that gesture could improve communicative success over verbalizations, but gesture alone was not more successful than the mixed condition. Fay concluded that gestures were used to bootstrap vocal communication. Fay et al's study evidences the importance of gesture, which is the means of communication for this study.

Most recently, the Language Evolution and Cognition lab at the University of Edinburgh has run a pilot study on a gestural communication task using the iterated learning model. The unpublished study headed by Kirby and Smith sought to determine when we might recognize language through investigating the mechanisms that drive change in the system and the development of systematicity. In the experiment, participants (5 chains of 5 generations) watched video stimuli of a ball moving in one of four pre-determined manners and paths, and after viewing the stimuli, the participants recorded themselves gesturing the target video (with an 8 second limit). Subsequent "generation" participants would then view the previous participants' gesture video, using it to guess which video in an array matched the gesture. This generation would similarly record themselves gesturing the target stimuli, and the process would be continued by the next generations. The researchers discovered the gestural communication system developed systematicity over the generations, becoming compositional (i.e. manner and

path were gestured sequentially, as different pieces of meaning). Also, the system moved from pantomime to symbolic gestures, starting from ball-shaped gestures to pointed fingers. The current study takes up this line of research, extending it to investigate the how dyadic interaction, allowing for both horizontal and vertical transmission, may affect the communication system and how it is transmitted over generations.

Research Questions and Hypotheses

This study extends previous studies on the evolution of communication systems in the laboratory, but crucially employs an interactional paradigm in which Director-Matcher (akin to speaker-hearer) dyads participate in a gestural communication task in each other's presence, and with an observer who is part of the next generation of interacters. This study will addresses the following research questions:

1. How do communication systems evolve in dyadic interactions, and how might interactional features of the dyad facilitate the evolution/transmission of the system?
2. Which features are present in lab experiments involving interactive communication games? And, what are the function of these features?

Given the findings of previous studies, and our intuitions about dyadic interactions, we hypothesized that over generations within a chain: 1) gestures should become shorter, moving from iconic pantomime to symbolic; 2) guess time should decrease, while accuracy increases; 3) gesture should become less variable. We investigated the role of interactional features as they arise in the data, and hypothesized that some features, such as eye gaze, may facilitate the evolution of the communication system.

METHODOLOGY

Based on prior work within gestural communication, and building from a pilot study using similar methods, a dyadic gestural communication task was carried out at the University of Edinburgh's Language Evolution and Cognition lab. Following the iterated learning paradigm, participants in this study were asked to communicate, in pairs, a series of meanings using only gestures, which would be passed on to the next generation of users via a non-interacting observer. The goals of this experiment were to observe the evolution of the communication system, hypothesizing a shift from iconic, pantomime gestures to more abstract (symbolic) ones, as well as to determine how interactional features might function in such a developing system.

Participants

The participants were recruited from a university employment website and were restricted to right-handed, native English speakers with no knowledge of signed languages. All participants were reimbursed for their time. Four chains of five generations were run, wherein the most experienced member of the interacting dyad would leave in the next generation to allow the observer to become part of the interacting dyad and a new observer would be added (as in Figure 3.1).
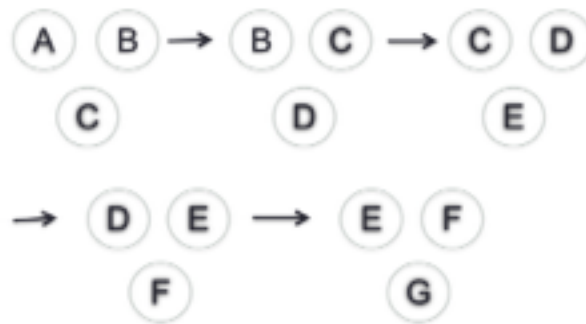
Figure 3.1: Participant Chain Structure by Generation

In the first generation, the first participant to sign up (participant A) was removed for the next generation. Also, the last observer, who would not become part of the interacting dyad, would perform gestures for the series of meanings to a camera without an interlocutor present. Two chains were assigned to a time-pressure condition, while the other two chains were not.

Stimuli

The meaning space for this study was borrowed from a previous, non-interaction-based pilot. The moving stimuli were of a ball moving along a particular path and in a particular manner (see Figure 3.2, for a still image example). Combining the four manners (roll, bounce, spin, jitter) and four paths (flat, slope, S-shape, circle) yielded a total of sixteen manner-path meanings.
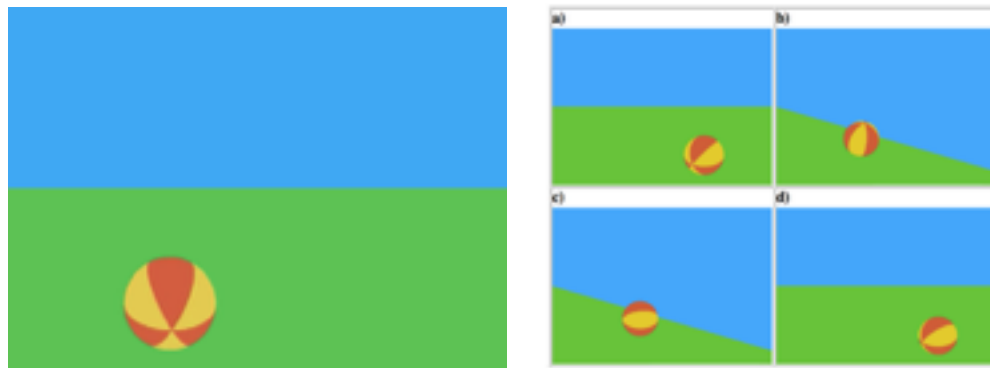
Figure 3.2: Stimuli (Still Frame) Sample: On the right, the Director's target stimulus (here, Roll-Circle); on the left, the Matcher's options (here, a) Spin-Circle, b) Spin-Slope, c) Roll-Slope, d) Roll-Circle))

Each participant gestured each of the meanings once during a generation. These stimuli were selected for their ability to be gestured either simultaneously or sequentially, as a means of investigating the level of compositionality demonstrated in the developing system. These moving stimuli were viewed on touch-controlled Apple iPad Minis, positioned in view of the participants, but so as to not act as a barrier between the dyad. The stimuli were displayed on the iPad screens through gifs (looped videos) and are available for viewing online (Director: *http://*

Procedure

Participants played a gestural communication game similar to charades, albeit with limited meanings. Three roles were taken up by the participants: Director, Matcher, and Observer. All three participants, as well as an experimenter, were present in the experiment setting. Directors (D) and Matchers (M) were video-recorded, and sat diagonal to one another, with an Observer (O) and Experimenter (E) on either end of the table (as seen in Figure 3.3). iPads were placed within view of the participants, but not in a manner as to obstruct the camera view or the interaction.
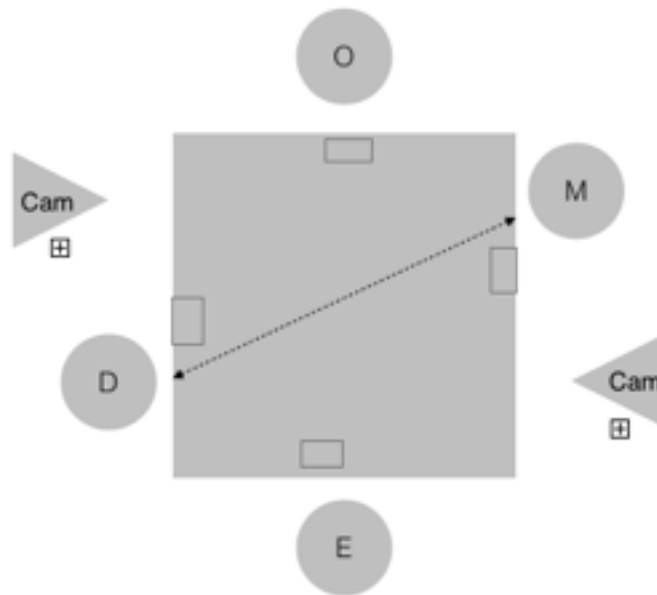
Figure 3.3: Experimental Set-Up

Directors gestured the moving image they had viewed on the iPad screen, Matchers provided a guess as to which meaning matched the Director's gesture, and Observers watched

57

the interacting dyad. Directors controlled the viewing of the iPad images, but were monitored by the Experimenter; Directors could, however, watch the stimulus as long as necessary to understand the meaning. While the Director gestured, the iPad screen would not display the moving stimulus to prevent mirrored gestures. While Directors viewed the target meaning, Matchers viewed an array of four moving images. The array contained the image with the target manner and path, one with the target manner and false path, one with the target path and false manner, and another with the combined false manner and path. The false manner and path was randomly generated. Once the Director gestured the target image, the Matcher would verbally provide the letter (A-D) that corresponded with the selected image. An Experimenter would give feedback about the selection: Correct, or Incorrect plus the correct option. Throughout a generation, the interacting dyad would switch between the Matcher and Director roles every four rounds, for a total of 32 rounds. The Observer in a given generation remained in that role, though they were told they would participate more actively in the next generation. Observers also used iPads to view the same arrays as the Matcher, had visual access to the Directors' gestures, and access to the Experimenter's feedback. All participants controlled their assigned iPad, but were guided and monitored by the Experimenter. The Experimenter also recorded the accuracy of Matcher guesses.

The two chains under the time-pressure condition were timed for each generation. Participants were told the dyad with the quickest time and highest accuracy (each inaccurate guesses added two seconds to the total time) for their chain would receive extra monetary compensation. At the start of each generations' turn, the participants were told the time to beat from previous generations. The timed condition was introduced as an external pressure on

communication, which we hypothesize would lead to the development of a more abstract communication system more rapidly than untimed chains.

Coding

Audio-video recording of the directors and matchers was done with Logitec wide-angle cameras (positioned in measured locations across from interacting participants), and video was fed to Apple MacBook Pro computers for storing. The video-collected data was coded for handshape, gesture start and stop times, guess accuracy, guess time, and discourse elements such as repair and eye gaze. Time stamps were recorded for gesture start and stop times, eye gaze, and guess times using the video software Inqscribe. Handshape was coded with a number and letter assignment adapted from the Hamburg Sign Language Notation System, and considered elements such as orientation (z or y axis as relative to the Director), movement (e.g. roll, circle, bounce, yaw), handshape (for both left and right hand, if necessary), open or closed fingers (for the ball), and the use of the forearm (as a horizon line). Eye gaze was coded with a letter indicating the focus of the gaze, including specific focals such as participant (P), gesture (G), screen (S), experiment (E), and observer (O), as well as non-specific focals, namely directions (e.g. up, down, left, right, across). Time stamps for shifts in eye gaze were also recorded, wherein the time was recorded once the gaze had reached its focal point, but the time of actually shifting remained coded under the previous focal point duration. Finally, repair sequences were flagged for further qualitative analysis, indicating whether they might be self-initiated or other-initiated repair.

RESULTS

Here we present the results, including accuracy over generations, timing for gestures and guesses, handshape structure, and the most prevalent features of interaction, eye gaze and repair.

Accuracy

The target video selection accuracy of Matcher guesses, overall, improved over generations. On average, the accuracy of the untimed conditions was higher than the timed conditions. Disruptions in accuracy in certain generations occurred if an unexpected delay arose between generations.

Timing: Gestures and Guesses

On average (across chains), the time per generation decreased. Exceptions, however, did occur when longer breaks occurred between generations due to unexpected delays in the the experiment. Timed condition generations were quicker overall, compared with untimed condition generations. The average gesture length (in seconds), as seen in Figure 3.4, decreased over generations (averaged across all chains), wherein there was a marked drop after the first generation, and a minor decrease in subsequent generations.
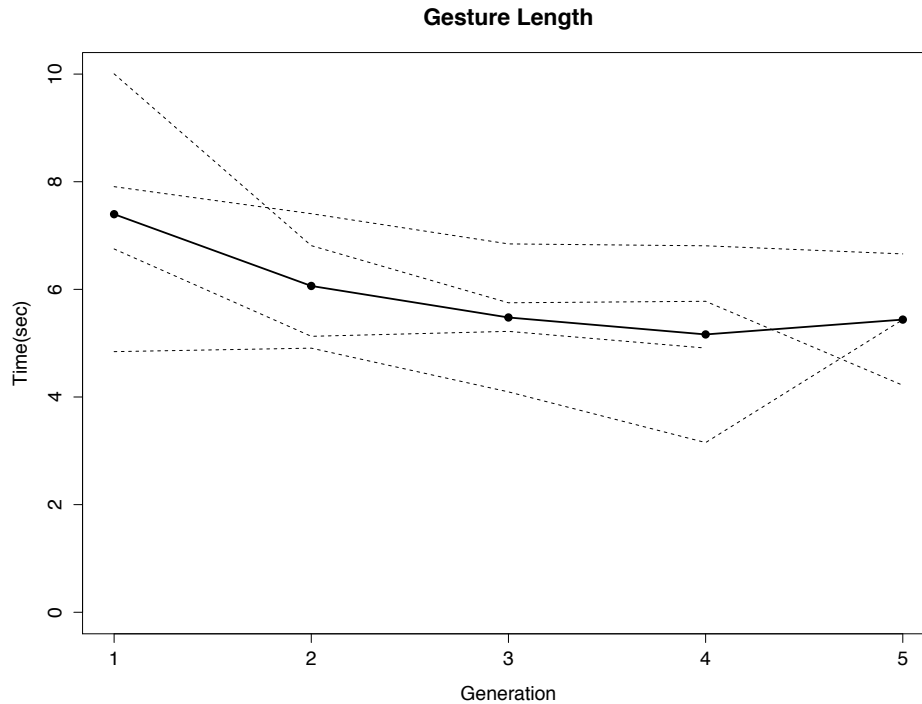
Figure 3.4: Average Gesture Length For Chains: The bold line indicates the average gesture length across chains, by generation. The dotted lines show the average gesture length for a given chain, by generation.

On average, the time to guess the gesture-video match shortens over generations (see Figure 3.5 below), while the accuracy of guesses tends to increase. Slight disturbances occur when delays between generations occur, suggesting a possible memory constraint as well. Analyzing each chain individually, we found that while patterns of decreased time of gestures and guess time occurred universally, there were more nuanced patterns about when guesses were provided during the gesture sequence.
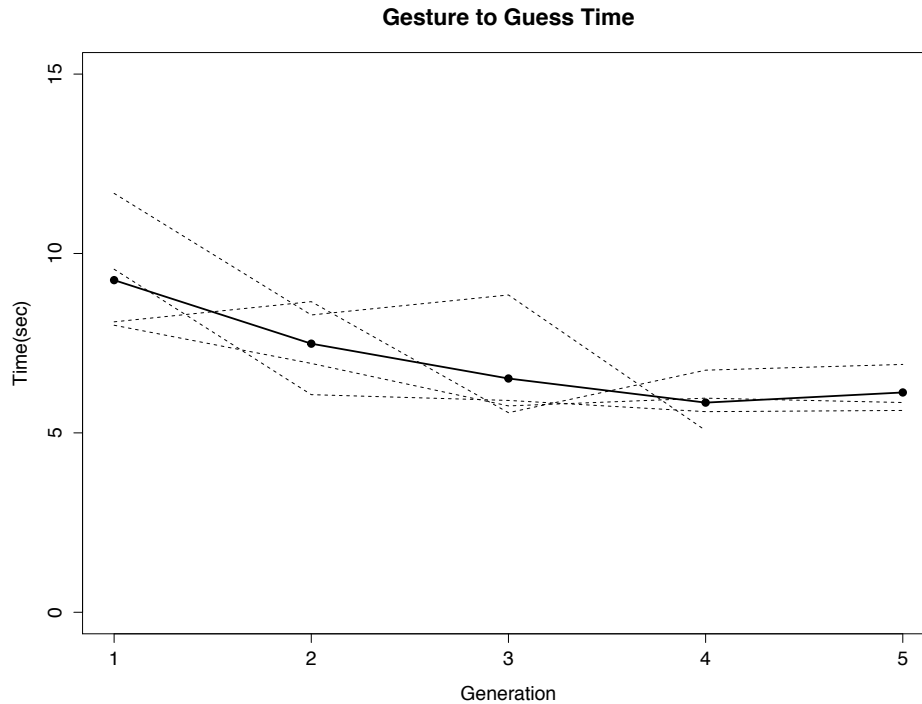
Figure 3.5: Average Guess Time For Chains: The bold line indicates the average guess time across chains, by generation. The dotted lines show the average guess time for a given chain, by generation. Timing started with the onset of the Director's gesture, and ended with the Matcher's guess.

In the untimed Chain 1 (Figure 3.6), the first generation provided guess immediately following the end of the gesture, however, over generations, guesses would occur during the gesture sequence. In contrast, Chain 2 (Figure 3.7) generations demonstrated a nearly 5 second gap between the end of the gesture and the Matcher's guess. This could have been a result of the Chain's tendency to not repeat the gestures. Although, by the last generation in Chain 2, gesture length and guess time had converged.

Chain 1: Generation Averages
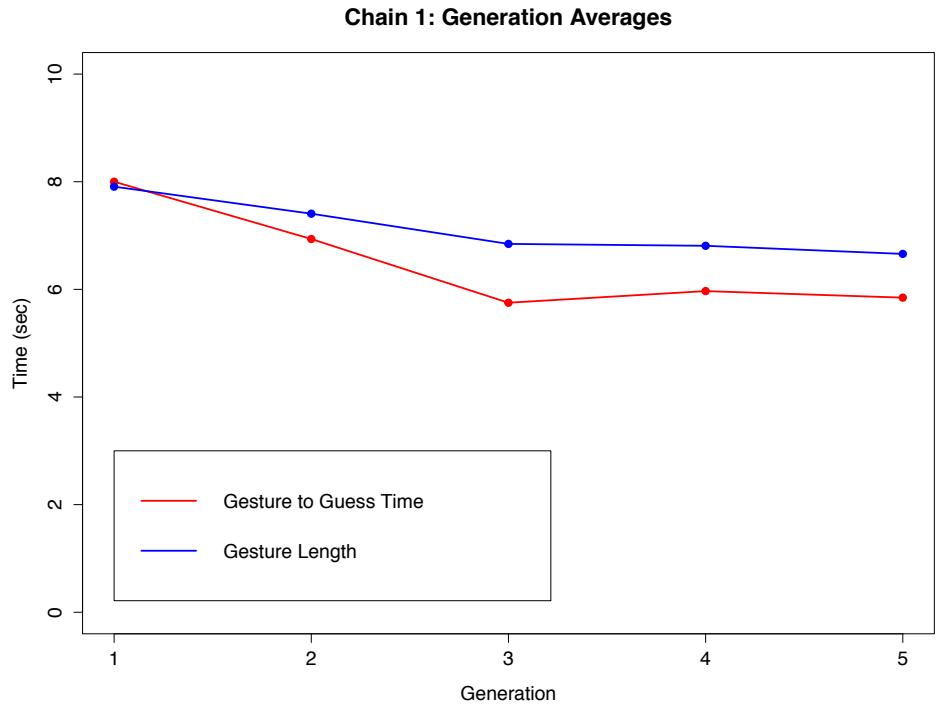
Figure 3.6: Chain 1 Averages: Generation 3 exhibits more marked drop in guess time than other generations.

Chain 2: Generation Averages
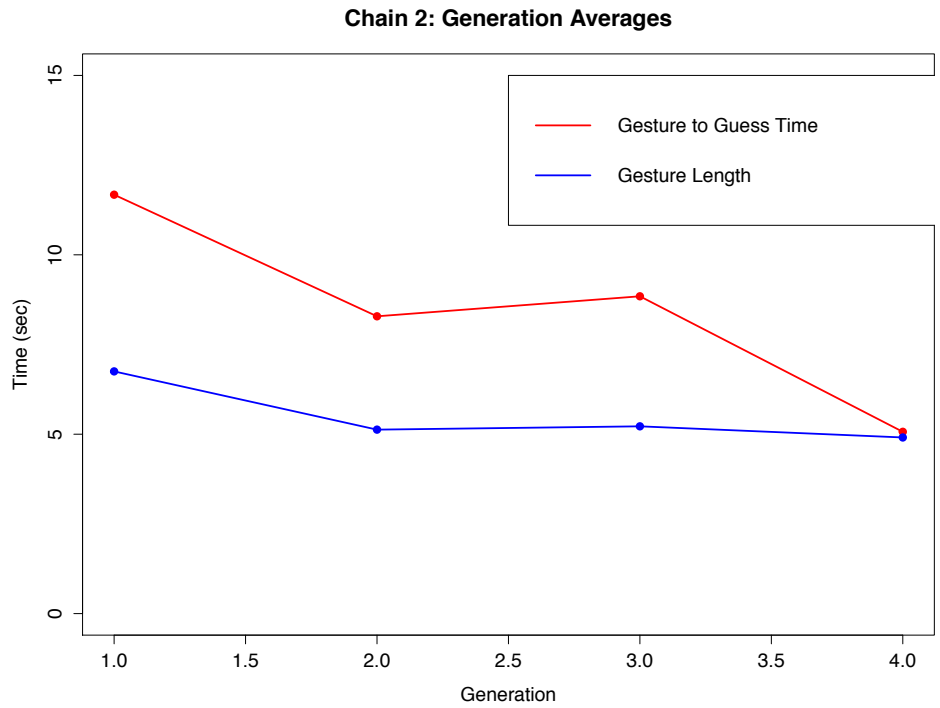
Figure 3.7: Chain 2 Averages: Guess time decreases overall, most notably in generations 2 and 4. Gesture length decreases slightly overall.

For the timed conditions, Chains 3 and 4 exhibited variable strategies as well, suggesting lineage-specific patterns of interaction that might not be affected by time pressures but rather communicative ones. Chain 3 (Figure 3.8) presents a similar pattern to the untimed condition in Chain 2 wherein guesses took place approximately 3 seconds after the Director's gesture had ended. While the time constraint could be a motivator to keep gestures short and unrepeated, the same cannot be attributed to the pattern found in Chain 2. The preference not to repeat gestures appears to be a choice for interaction and not as a result of the time pressure introduced in later chains. Conversely, Chain 4 generations (Figure 3.9) displayed little difference between the gesture length and guess time, with guesses initially occurring during the gesture- which ended nearly immediately after the guess was provided. By the last generation, gesture lengths had shorted enough to result in guesses being provided after the gesture had ended. This marked difference in guess time and gesture length preference within the same condition could indicate that patterns of interaction- as typically determined in the initial generation- will be maintained throughout the chain, regardless of external pressures.
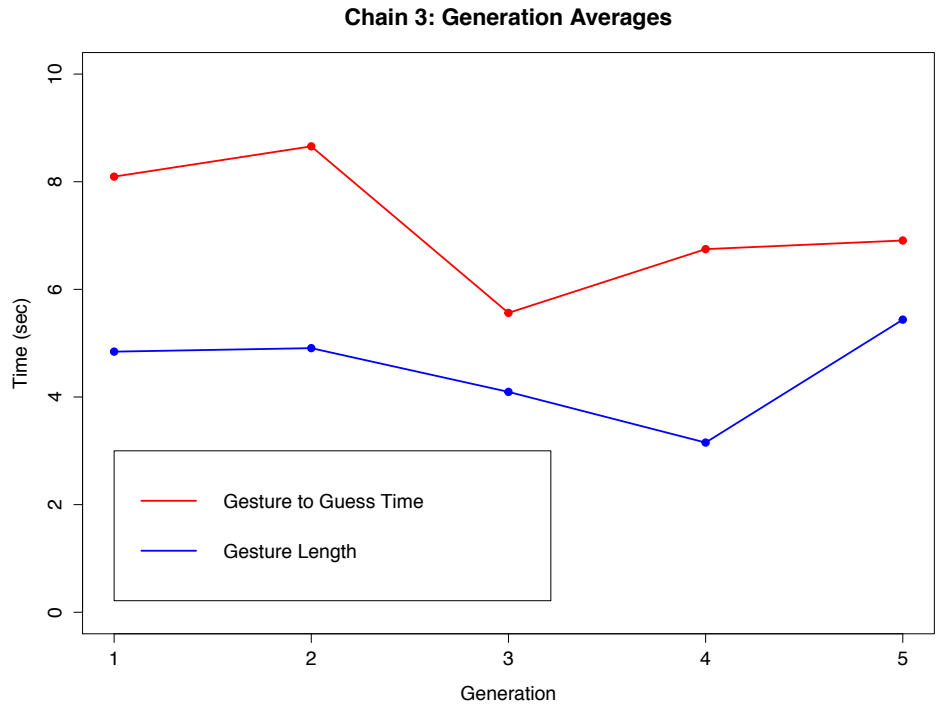
Figure 3.8: Chain 3 Averages: Guess time decreased from generation 1 to 5, but not in a steady manner. Gesture length decreased until generation 4.
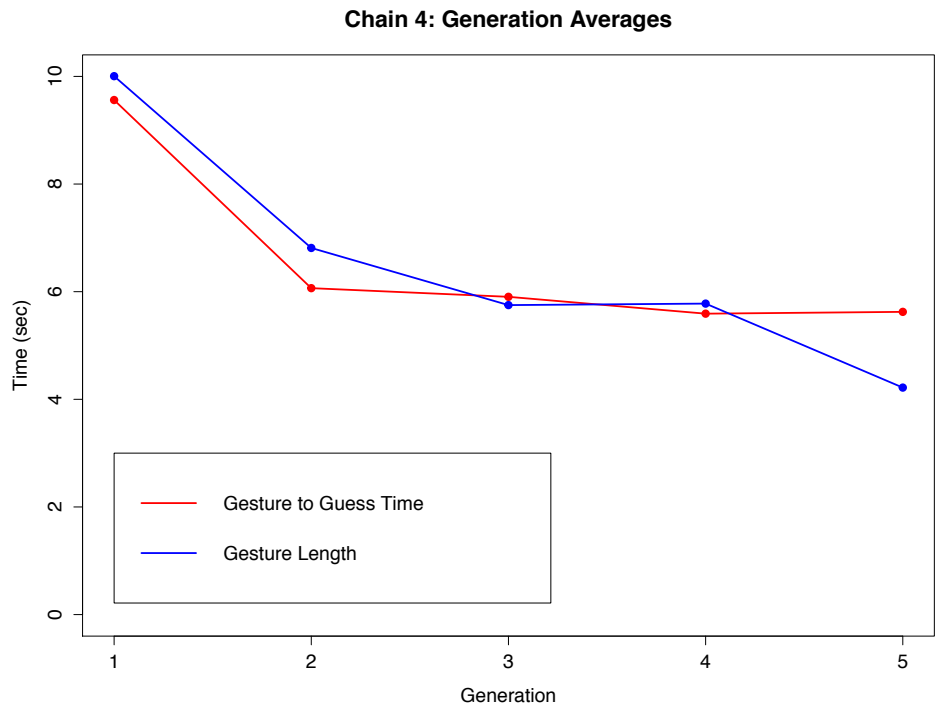
Figure 3.9: Chain 4 Averages: Gesture length and guess time decreased dramatically after generation 1. Gesture length decreased again at generation 5.

Overall, Chains 2 and 4 had the greatest decrease in gesture start to guess time (all chains showed decreases), dropping from 12 seconds to 5 seconds and 10 seconds to 6 seconds, respectively. Chain 2 may have exhibited a more dramatic change due to the slow nature of the first generation participants. Chain 4 also had the greatest decrease in gesture length over generations, starting at 10 seconds and ending with 4 second long gestures by the final generation. All other chains maintained a 2 second differential in gesture length, which was typically achieved gradually over generations, expect in the timed conditions, which did display more marked decreases after certain generations.

Structure and Handshape

Participants overwhelmingly produced gestures for manner and path simultaneously (81% of all gestures) as in Image 1, with few produced sequentially as in Image 2 (e.g. manner followed by path). The participant in Image 3.1 is performing the Bounce-Flat target form. To simultaneously represent the manner (Bounce) and path (Flat), the participant holds her moving hand/arm in a fist shape to represent the ball, while the static arm is positioned as a flat horizon line. The balled fist moves up and down to indicate a bounce motion, while simultaneously inching from the participant's left to right on a flat trajectory. The participant, in the Director role here, maintain eye contact with the Matcher throughout her performance.
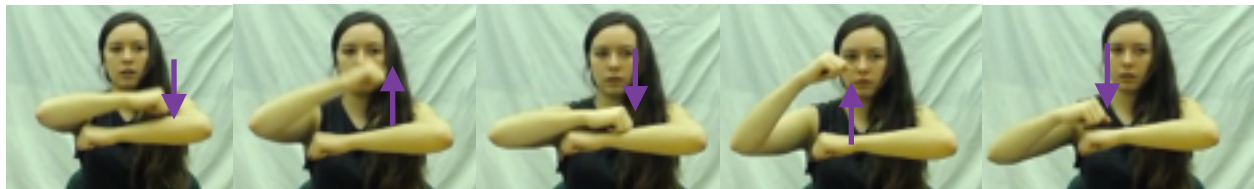
Image 3.1: Simultaneous Structure (Bounce-Flat)

The participant in Image 3.2 demonstrates a sequential gesture, first performing the target manner (Spin) and after its completion, gestures the target path (S-shape) across the table's surface. The ball is represented by both hands, spread and open, as if encompassing the ball between them. This representation only occurs during the manner gesture, while no ball representation is used in the path gesture. Rather, a pointed finger draws the path along the table. To repeat the ball representation might be redundant, and therefore not well suited for the sequential gesture structure. Through most of the gesture sequence the Director maintains eye contact with the Matcher, though in the third image she looks down to her own gesture, now performed on the table, potentially drawing her co-participants gaze there as well (given first part of the gesture occurred in air, above the table).
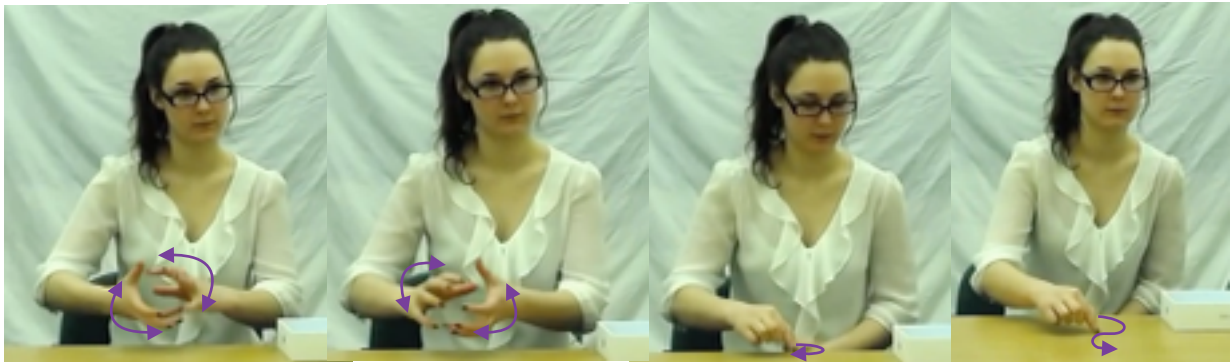
Image 3.2: Sequential Structure (Spin-S-shape)

Nearly all manner and path gestures were distinct, however some generations produced identical gestures for the manners roll and spin, as with the participant in Image 3.3. The gesture for both roll and spin, for the participants in this chain, was hands held statically in a ball shape, and moving in the target path trajectory. In the example below, the gesture could be interpreted as either Spin-Slope or Roll=Slope, wherein the path "Slope" is gestured twice, first as a initially diagonal horizon line, then simultaneously with the manner gesture.

67
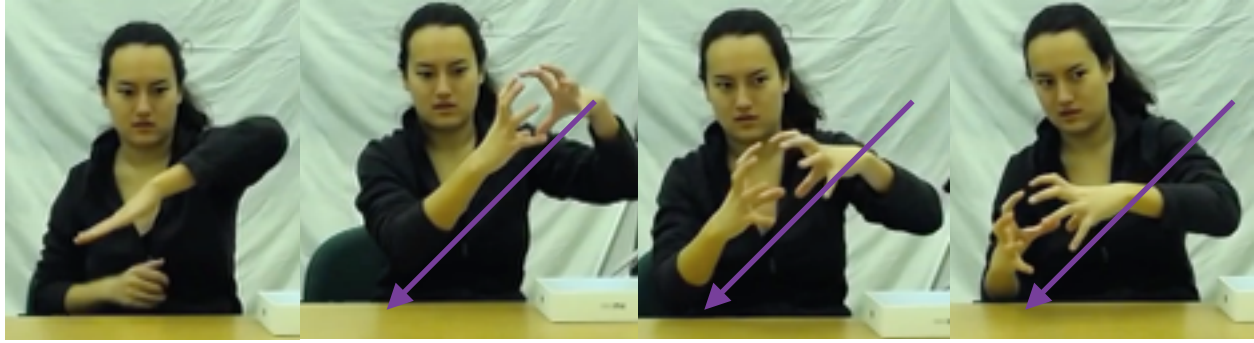
Image 3.3: Context-Dependent Gestures (Roll and Spin Manners)

Producing the same gesture for multiple meanings requires the Matcher to rely on the context of their array when guessing which meaning is being communicated. This strategy, while symbolically efficient, could led to communication errors.

Over generations, the orientation, movement, and handshape decreased in entropy; that is, the variation for particular meaning-gesture pairs within a chain dropped over generations- the participants conformed to certain gestures for a given meaning. The orientation tended to be on the y-axis (i.e. gestured in front of the Director, from left to right, for example), relative to the Director, which was likely optimal for Matcher observation. Handshape and movement were dependent on the meaning at hand, though they came to not differ within meanings for different participants in a chain. While early generations still navigate the communicative boundaries of their interaction and test new gestural strategies, participants eventually settle on a set of gestures that become conventionalized. Through this process, the variation and diversity of gestures decreases over generations as forms become systematic.

Director Eye Gaze

Eye gaze patterns of the Directors across generations are similar in terms of focal points, but different chains produce varying patterns that give precedence to one focal point over another. In addition to the frequency of certain focal gazes, measurements for the average time

(duration) Directors looked at focal points were also taken (graphs for average gaze duration for generations within a chain are in Appendix B).

Chain 1 Directors (Figure 3.10) more frequently gazed at their partner (i.e. the Matcher) than any other focal point; other glances that largely populated the Director's gaze were the iPad screen and their own gesture, occurring at a similar frequency. However, the duration of the eye gaze favored the iPad screen, followed by the participant Matcher, and then the Director's own gesture. The pattern that emerged in this chain was to start the gesture sequence by looking at the participant, then to look at their own gesture, gaze up at the participant near the close of the gesture sequence, and then to look to the screen.
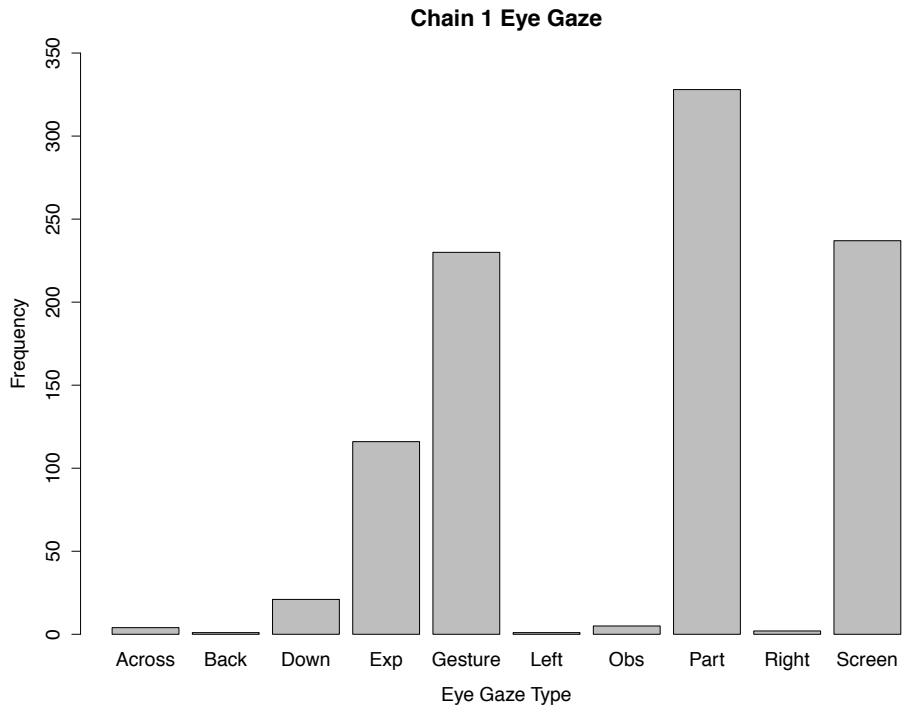
Figure 3.10: Chain 1 Eye Gaze: Frequency of looks toward focal points, including nondescript points (Across, Back/Behind, Down (to table), Left, Right), and specific points (Experimenter (Exp), Gesture, Observer (Obs), Participant (Part), and Screen)

Chain 2 Directors (Figure 3.11), on the other hand, looked more frequently and longer at the iPad screen, followed by the participant and their gesture. This may have been a result of these Directors not changing from the target stimuli slide to the blank gesture slide, making the screen a more salient focal point than in other chains. While participants received more glances, the Directors looked longer at their own gestures; again, this might be an effect of the target stimuli being present, and Directors attempting to mirror or copy the video. The divergence is also apparent in the eye gaze pattern common to Chain 2, which had Directors looking at the screen at the start of their gesture, then looking to their gesture, and glancing at the Matcher briefly before returning their gaze to the screen as the Matcher provided a guess.
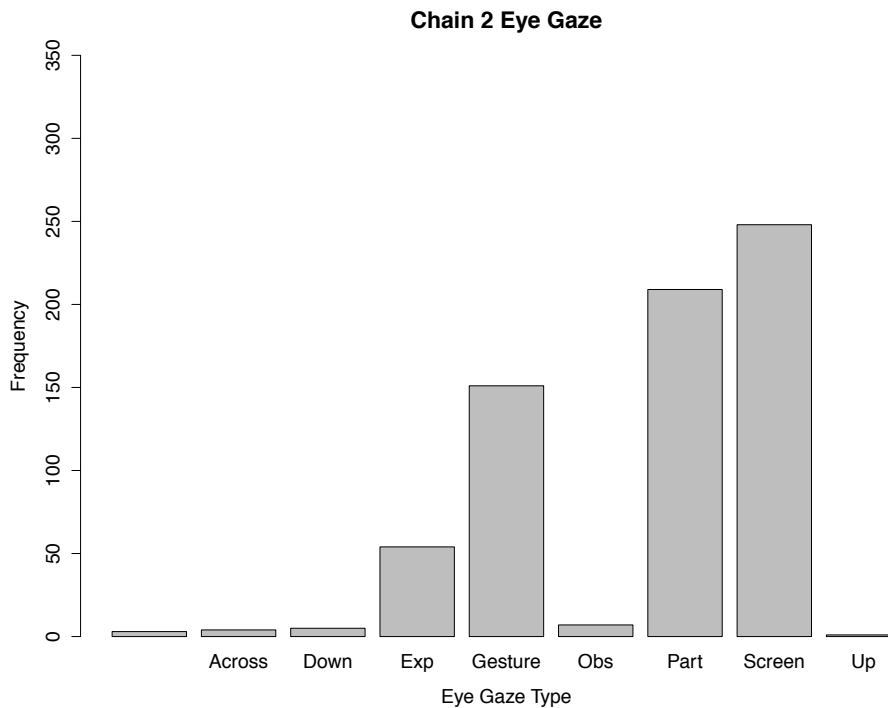
Figure 3.11: Chain 2 Eye Gaze: Frequency of looks toward focal points

Directors in the timed conditioned Chains 3 and 4 displayed similar frequencies and durations with their eye gaze. In Chain 3 (Figure 3.12), while Matcher participants received the most glances (followed by the screen and gestures), Directors maintained their gaze toward the screen and their own gestures for longer periods of time. The pattern of Chain 3 was similar to that of Chain 2: moving from the screen to their gesture to the Matcher and finally back to the screen.
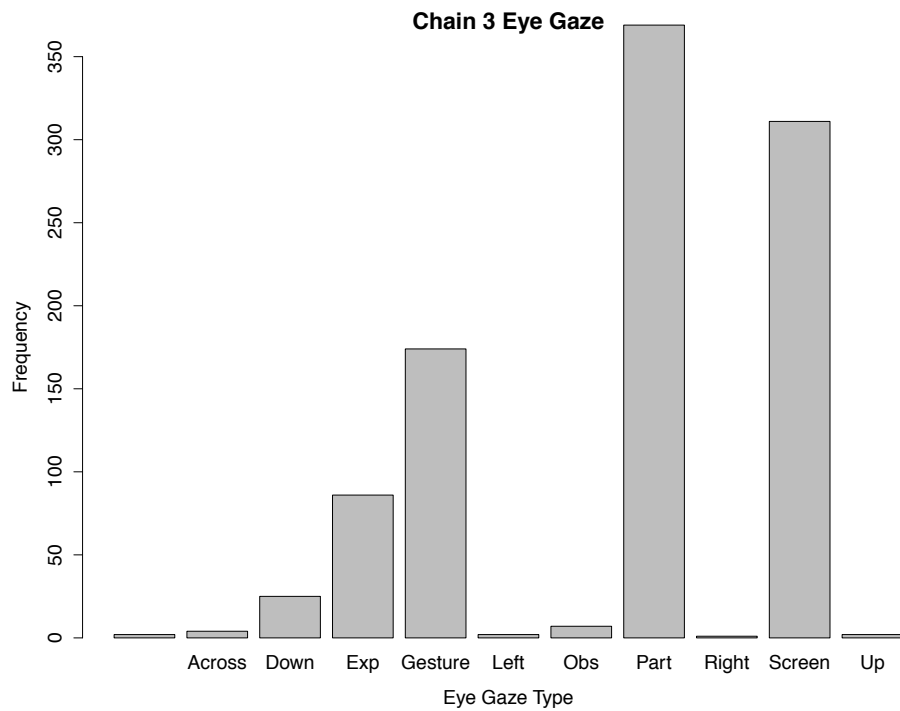
Figure 3.12: Chain 3 Eye Gaze: Frequency of looks toward focal points
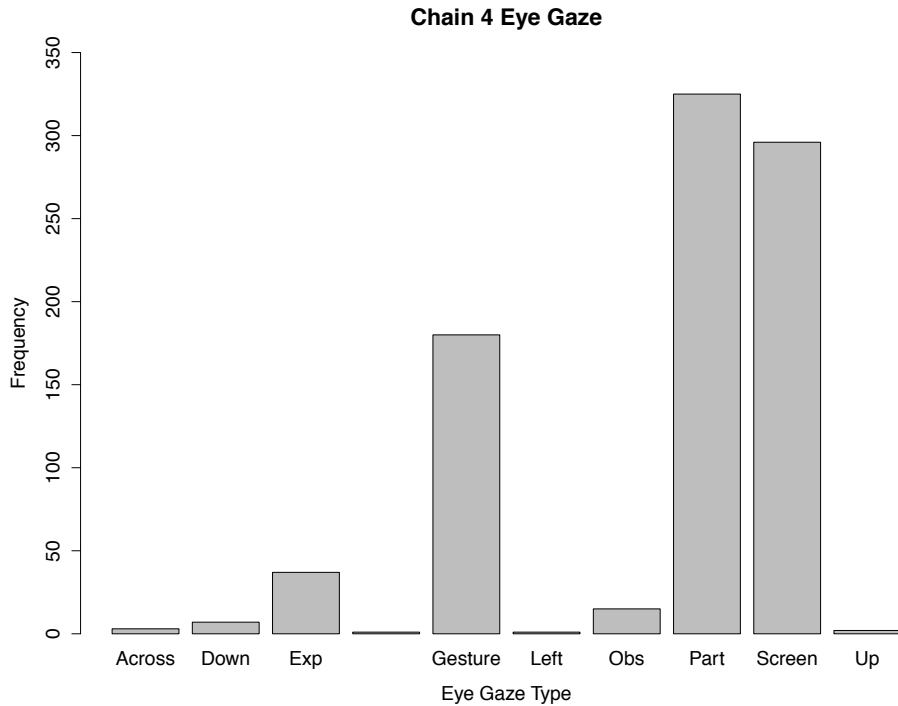
Figure 3.13: Chain 4 Eye Gaze: Frequency of looks toward focal points

Chain 4 Directors (Figure 3.13) had the same frequency and duration patterns for the top focal points, but these Directors carried out a different sequence for their changing eye gaze. Unlike the screen-initial eye gaze of Chains 2 and 3, Chain 4 Directors started their gesture sequences with looks toward the Matcher participant, followed by their own gesture, and finally ending with the screen (typically to change to the next stimulus, whereas Chain 2 and 3 Directors looked at the screen waiting for Matcher guesses).

Overall, Chains 1, 3, 4 had the most looks to the participant, though the timed Chains 3 and 4 had more looks to the screen than Chain 1. The untimed conditions had similar look frequencies to the screen, but were less than the screen look counts from the timed conditions. The timed conditioned Directors looked to their own gestures more frequently than the Directors in the untimed conditions.

72

Repair

Though repair did not occur in great frequency, there is a sufficient number of examples of the discourse phenomenon that lend themselves to analysis. Overall, the chains in this experiment yielded 36 instances of modification repairs (clarification and reformulation) and 73 instances of repetition repair. However, repair strategies occur in disproportionate amounts in different generations; that is, all repair strategies decrease over generations (Figure 13.4), though modification ones do so more drastically (even significantly, as demonstrated by a linear regression model, predicting repair by generation: $p < 0.055$).
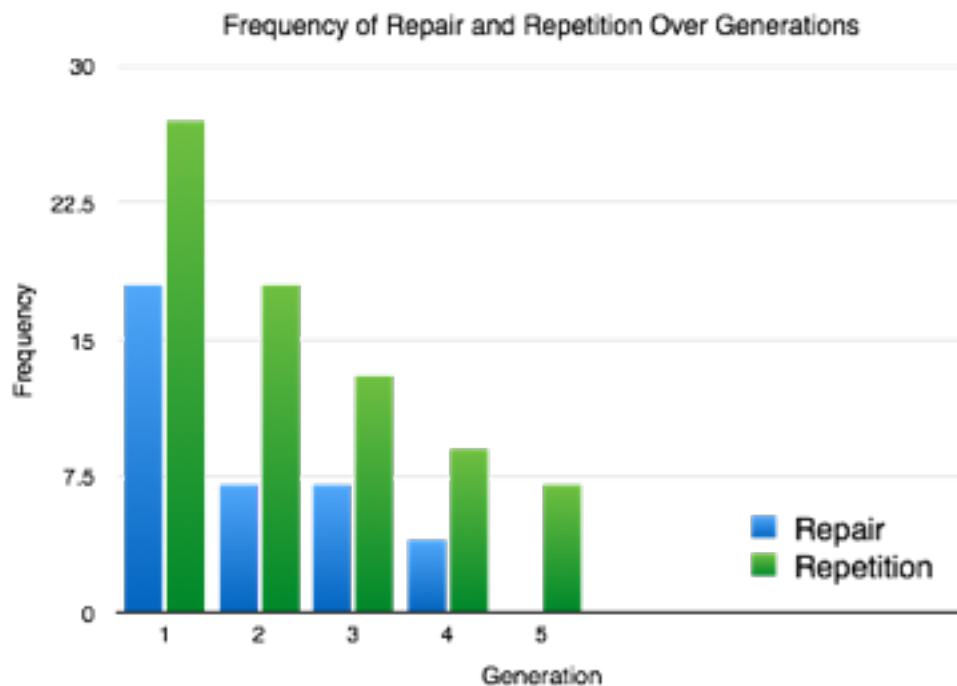
Figure 3.14: Repair and Repetition over generations: Repair (clarifications and reformulations) decrease more dramatically over generations than repetition, which still exhibits less frequency over generations

Variation also exists in the amount of modification repair individual chains (lineages) preferred to use. The untimed chains 1 and 2 performed 8 and 3 repairs, respectively; timed chain 3 did not produce any repairs, though timed chain 4 produced 25. Chains 1 and 4 exhibited the most repair,

which indicates time pressure may not be a factor in participants' decision to make use of repair. The variation between chains can be seen in the red dotted-lines in Figure 3.15, though the average (indicated by the bold line) of repair use for the generations within all the chains does demonstrate a significant relationship ($p < .05$) between generation and repair.

Average Repair per Generation
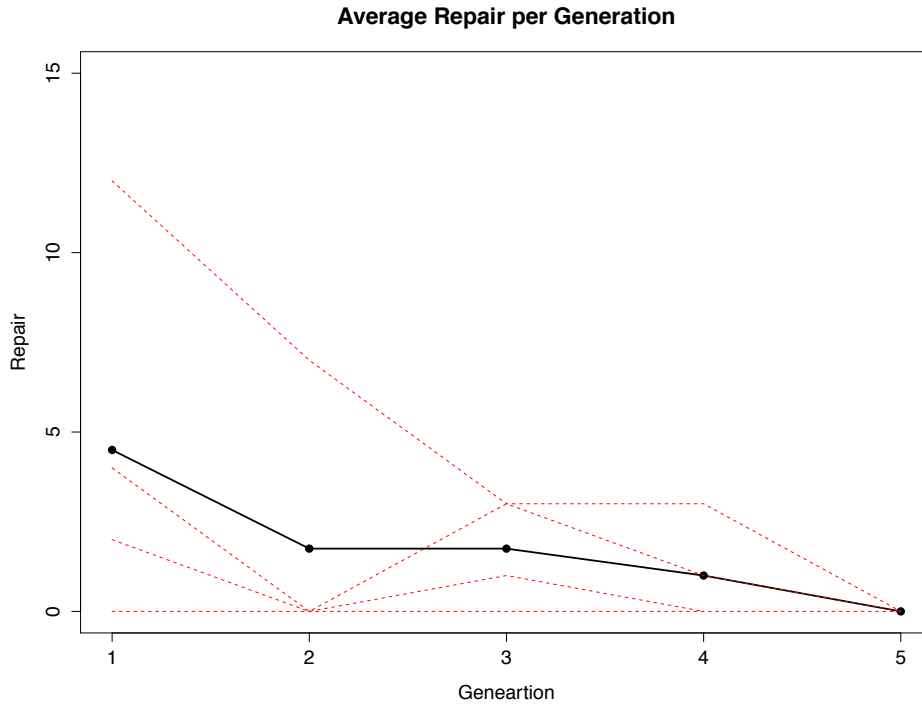
Figure 3.15: Modification Repairs across Generations: Variation exists between chains, but on average the amount of repair per generation is significant and existed a downward trend

Repair may have been constrained or limited due to the nature of the experiment; participants may have felt unable to produce repairs given the constraints of gestural interaction, the experimental setting, or the task itself.

In the first sequence of the Director's gesture in Image 3.4 is a new gestural strategy introduced by the Director in this early round of the second generation. Earlier gestures had Directors using their fist to represent the ball and its movement. Here, the new gesture requires the use of the body to describe the movement, while the arms are held wide to embody the ball.

74
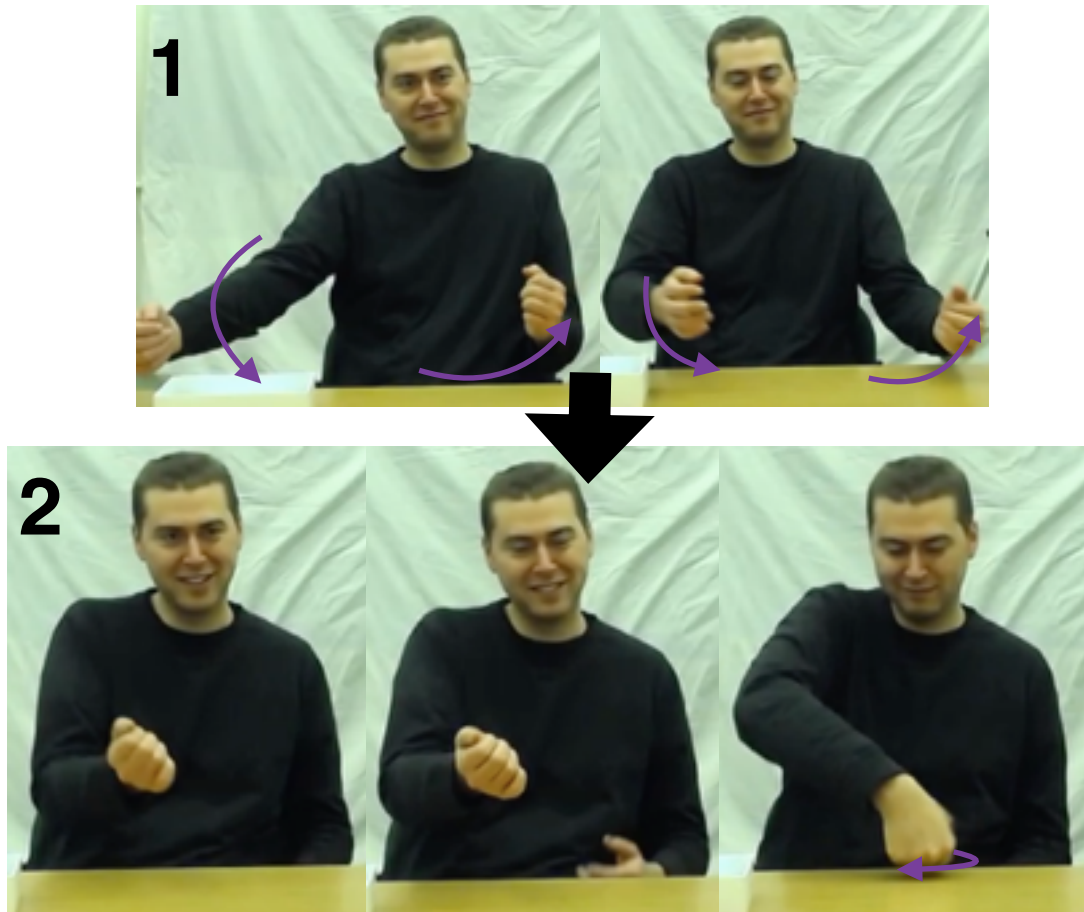
Image 3.4: Reformulation Repair: Trouble source is new gesture type; it is repaired to the conventionalized gesture

The second sequence in Image 3.4 is the repair sequence, in which the Director returns to the conventionalized gesture (in place from the first generation) after there is no uptake from the new gesture structure. The conventionalized gesture is performed with the Director's fist moving to convey both manner and path simultaneously along the flat surface of the table.

A sample clarification repair is performed by the participant in Image 3.5, gesturing Roll-Slope. The clarification is from an initial sequential path (Slope, indicated with the diagonal forearm) then manner (Roll, conveyed through the use of both hands forming a ball shape and rotating one hand over the other) gesture to a simultaneous one. The original path gesture is omitted in the repair, but is replaced with the continuous downward diagonal movement,

combined with the same manner gestures as before. The simultaneous gesture emphasizes the manner and path, since both are being performed continuously throughout the repaired gesture.
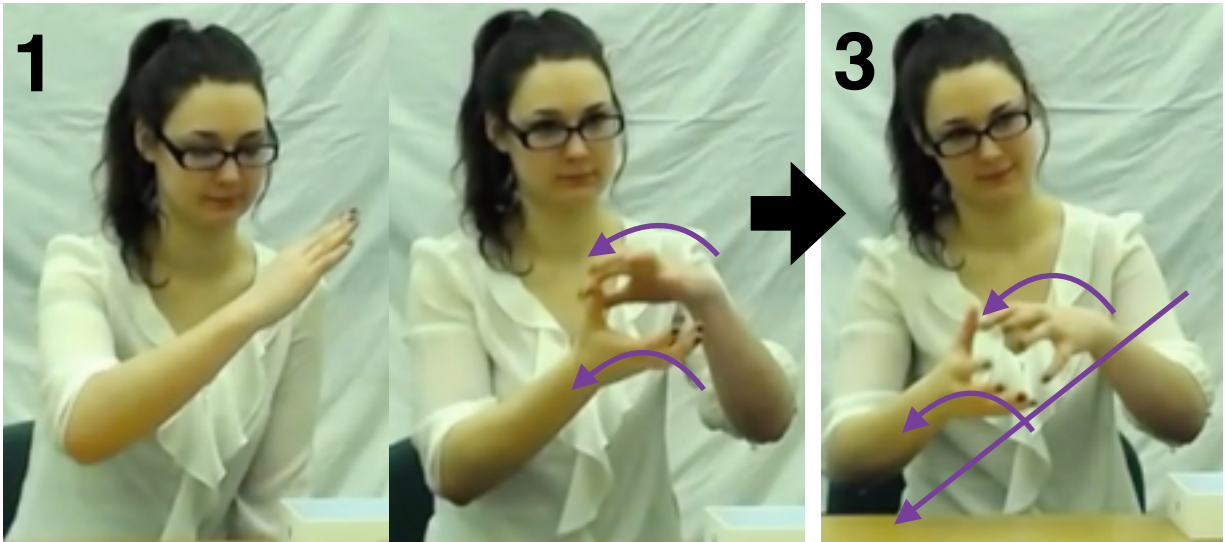
Image 3.5: Clarification Repair: The initial, sequential (1. Slope path, 2. Roll manner) gesture is repaired to a simultaneous one (3), wherein the manner and path are emphasized with a continuous rolling movement downward

Clarification repairs typically included emphasis of the repeated form, which might include either manner or path, or a combination of the two. The emphasis is paired with eye gaze as well, which either monitors the Matcher's comprehension, or directs attention to the specific point of repair.

It is hypothesized that the eye gaze patterns (and potential facial expressions) of Matchers would be indicators of confusion or misunderstanding- particularly in conjunction with delayed guesses- that would inform the Director that repair must be done. Less often, repairs that were self-initiated by the Director tended to occur due to Director error (e.g. performing the wrong manner or path- to which the Matcher would not have epistemic access). Image 3.6 depicts the process of self-initiated self-repair (Jefferson, 1974) performed by the Director, who began to

gesture the previous target image and had to repair to the new target once he recognized the error.

Image 3.6: Self-Initiated Self-Repair

In instances of self-initiated repair from Directors, the repair was signaled by a particular eye gaze pattern and/or other facial gestures. For example, in the image above, the Director signals a "recall" event by looking upward. He then changes the gesture motion to the correct target, performing it with an apologetic smile. Participants seem to be bootstrapping interactional features common in their native language use to the novel system used in the experimental setting.

DISCUSSION

Evolution of the Communication System

Through the cultural and lineage-specific transmission of the gestural communication system, the gestural meanings became systematic and conventionalized. In an dyadic paradigm, participants are forced to negotiate not only the form of the communication, but also what those forms come to represent. This negotiation occurs over time, via simulated generations, until a conventionalized system is agreed upon through continuous use. Participants can identify when the system reaches this point by observing the communicative accuracy between the dyads- that

is, not only can the participating dyad note this but the observer can as well. Once a dyad demonstrates reliable communicative success, the observer who will participate in the next generation will be less likely to make changes to the system in progress, so as not to disrupt the advances of previous generations. Likewise, if early generations have not established a reliable and conventionalized system, observers can use the information from the dyadic interaction to introduce changes in the gestural system. The interacting dyad can propose changes to the developing system as well; however, if one participant feels there has been a contradiction to an emerging conventional system, they may be reluctant to take it up, or risk misunderstanding. The methodology used for this study appears to get at the crux of language evolution: how people interacting for a common goal develop and transmit a communication system that is effective.

There is evidence that the gestural communication system in this study did evolve through cultural transmission. Over generations, the gesture length (in terms of time) reduced, in some chains significantly. Shorter gesture length would lend users to using more abstract, symbolic gestures that did not represent the meanings in an iconic, or pantomime, manner. In addition, gestures produced for particular meanings became more similar over generations, namely in terms of handshape, hand movement, and orientation. The variation in gesture meanings was slight, as 81% of all gestures were simultaneous, grouped meanings. However, between chains, there was evidence for lineage-specific gestures. Different chains of generations therefore developed differing gestural systems. For instance, some chains consistently used the forearm as a redundant feature, while others drew horizon lines from the video with one finger. Some chains chose to use two hands for gestures, while others used one, and some preferred to repeat gestures while others did not. The lack of variation within chains, but diversity between

78

chains, demonstrates that while no one system is fit for all groups, they do conform to specific standards of an efficient communication system within each lineage.

One counter-example to the efficient system is the presence of some context-dependent gestures. In some generations, participants would perform the manners "roll" and "spin" identically (see Image 3.3 above). The homophonous-like gestures could only be distinguished from one another if they did not co-occur in the array presented to the Matcher (of which the Director did not have knowledge). In the absence of their co-presence, the Matcher would correctly guess the meaning, however if roll and spin were both options for the Matcher (as randomly determined with the code used to run the experiment), Matchers were faced with a 50% chance at a accurate guess. Given the relative infrequency in which the co-presence of "roll" and "spin" would occur, the strategy to limit the gestures overall is not a counterintuitive one. Natural languages have many homophones whose meaning can derived from context, and if not, they can be remedied quickly after a misunderstanding. If the participants employed more repair techniques, the confusion that could arise from the conflation of "spin" and "roll" would be mediated through clarification requests or another form of repair.

Repair as a Sign of Conventionality

Though it did not arise frequently, repair was a feature of interaction that could function as an expressivity constraint in the evolution of the gestural communication systems in this experiment, as well as in natural language interaction. Repair was most frequently initiated by the Matcher (equivalent to the "hearer" in most studies of repair)- though performed by the Director in keeping with the preference for self-repair. The delay in providing a guess, or in maintaining eye gaze (see the section below for more in depth analysis on eye gaze) with the

Director after the gesture has ended could be indications of misunderstanding, and therefore interpreted as repair initiation. When gestures are not congruous with the emerging system (e.g. when participants want to "try something new" or change the systems dramatically), Matchers are more likely to misunderstand the meaning the Director is attempting to convey. This misunderstanding may more likely led to instances of repair. In this way, repair keeps the system in check, drives conformity, and aligns the participants in interaction.

In the conversation analysis-based repair literature, repair comes from two trouble sources: a problem of understanding or a problem of hearing (in this case, seeing). Different trouble sources result in different strategies of repair, ranging from repeats to clarifications, to reformulations. This study found two repair strategies that arose from the most prevalent trouble source: problems with understanding (indexed by eye gaze and facial expressions, to be discussed below). One strategy was to repeat the initial gesture (fully or partially), while the other strategy required modification of the initial gesture. The latter strategy could be performed in one of two manners: clarification or reformulation. Directors performed reformulations when the gesture in question deviated from the conventional system (as in Image 3.4); the reformulation would have them return to the system already in place for that particular chain. In contrast, repairs performed as clarifications are often exaggerated (or prolonged) repetitions of the original gesture (as in Image 3.5), sometimes isolated to the part (i.e. manner or path) perceived to be problematic. These types of repair could help solidify the developing system, as they make gestures more distinct for particular meanings yet more salient for learnability.

As a system emerges, undergoes negotiation for form-meaning matches, and becomes "fixed" over generations, repair should be a notable feature, particularly early in a chain. Repair

allows participants opportunities to align to certain gestural forms for target meanings. Deviations from forms which they have begun to converge on can lead to repairs to keep the system conventional and standardized. Repair should become less frequent over generations, and used when conventions are not met or a gesture is not seen by a Matcher. In the latter context, repetitions should be more prevalent and may even persist in later generations. This is in fact the trend we see in the data here, in which all generation 5 repairs are performed by repetition. We suggest that the types of repair strategies participants rely on is in direct relation with the stage of evolution of the language system.

The Functions of Eye Gaze

Natural, face-to-face conversations employ many discourse-level features that facilitate and progress the interaction; one powerful feature is eye gaze. For decades researchers have examined the function of eye gaze in conversation. Kendon (1967) proposed it had a turn yielding function in which speakers would yield their turn to another by selecting them as the next speaker through eye gaze. Goodwin (1980) furthered the study of eye gaze, claiming it helped manage the flow of talk. Speakers aimed to obtain the gaze of the hearer to establish mutual gaze, reinforcing attention. Also, speakers would direct their gaze to relevant members of the conversation, for instance looking at an unknowing (i.e. low epistemic status) hearer to deliver news. More recently, Stivers and Rossano (2010) found gaze to be a response mobilizing strategy in which mutual gaze between a speaker and hearer would extend the sequence at hand, while speaker and hearer gaze withdrawal would bring the sequence to a close. Interlocutors do not need to share eye gaze with one another in conversation though. Tomasello's (2008) theory of joint attention only requires interlocutors to share mutual gaze with the same object, person, or

other focal point. This shared gaze helps interlocutors recognize their shared attentional focus, and they can then elaborate on and speak more abstractly about that object, without fear of misunderstanding. These theories inform the analysis of eye gaze in this experimental setting.

A novel, and interesting, observation from this study is that not only is there cultural transmission of the communication system, but also of the interactional features that accompany them. Lineage specific pattern of eye gaze emerged in the dyadic interactions and were reliably maintained over the generations. While we might expect individual differences to play a greater role in discourse features, it appears that with a semi-structured "conversation" style, a pattern does emerge and tends to hold over the generations in a chain. Eye gaze patterns do not have the same manifestations in all chains, and are likely highly dependent on the participants and their environment. The most common patterns were G-P-S and P-G-S; given the screen prompts the next sequence, it is reasonable that it occurs at the end of the pattern frequently. However, participants might establish either the gesture or the Matcher as the first point of focus in accordance with their goals: gaze to the participants ensures that the Director has the attention of the Matcher before gesturing, while gaze to the gesture could alert the Matcher to attend to the gesture, as in Tomasello's joint attention theory. The environment does matter in emerging eye patterns, and has the potential to shift patterns. For example, generations that did not change the target stimuli while directing would more frequently look at the screen, instead of the participant. Furthermore, the duration of looks toward the focal points are very similar across generations in a chain, as though participants mediate their gaze so as to not look at a given focal point longer than their co-participant. Their conformity in eye gaze duration and patterning might be evidence of systematicity in the interaction in general, which couples with that developing in the

communication. That is, both the discourse features (of eye gaze, at least) and the communication are transmitted culturally through generations in an iterated learning chain.

A communicative function of eye gaze, as well as facial expressions, found in this study is that of marking a lack in comprehension, or a repair initiator. As mentioned previously, eye gaze patterns are one cue Matchers could utilize to signal they have not fully grasped the meaning performed by the Director. In fact, deviations from eye gaze patterns, either in terms of focal point or duration, come to be recognized as cues for misunderstanding and the need for repair. For example, prolonged Matcher gaze to the screen, postponing the time to guess, for some "experienced" generations can signal miscommunication. "Experienced" generations have become entrained to the timing for guesses and looking at particular focal points, and can therefore be more susceptible to deviations from what has become standard within their lineage. In Image 3.7 a Matcher gazes at the Director's sequential gesture of "Roll-Flat"; manner (Roll) is performed first (Still 1), with both flat hands moving one over the other repeatedly, then path (Flat) is performed (Still 2) with a flat hand drawing a straight line in front of the Director.

Image 3.7: Stills 1-2: Initial Gesture: Matcher (left) and Director (right) from Chain 5, Gen 2

As the path gesture comes to an end, the Matcher has diverted her gaze to the screen, yet does not give a guess. Though an example from Chain 5- omitted from other data analyses due to its shorted generation length- these two participants often guessed before the gesture was completed (with an average gesture length of 8 seconds, and guess time of 6 seconds). The delay in guess time incited the Director to take up the gesture once more (Still 3), beginning with the repeated manner gesture (Still 4).
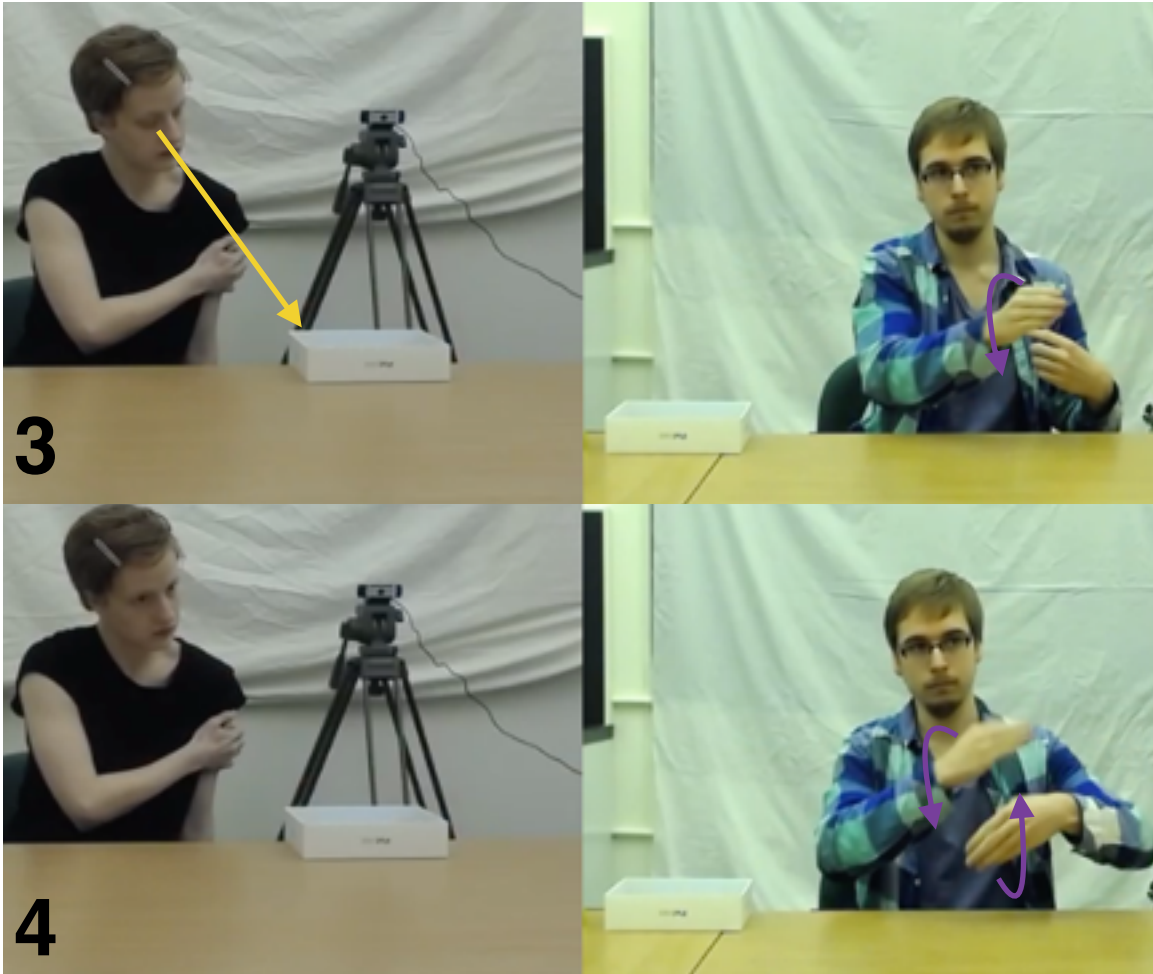
Image 3.7: Stills 3-4: Matcher Eye Gaze Eliciting Repair: Non-response and eye gaze prompt repair

However, the path gesture was clarified with a change in handshape (Still 5). Rather than a flat

hand moving along the flat path (conceivably confusable for a horizon line, sometimes included

in these gestures), the Director holds his hands spread- encompassing the ball- and move along

the flat trajectory. Once the clarification repair has been made, the Matcher returns her gaze to

the screen (Still 6), and provides a correct guess.
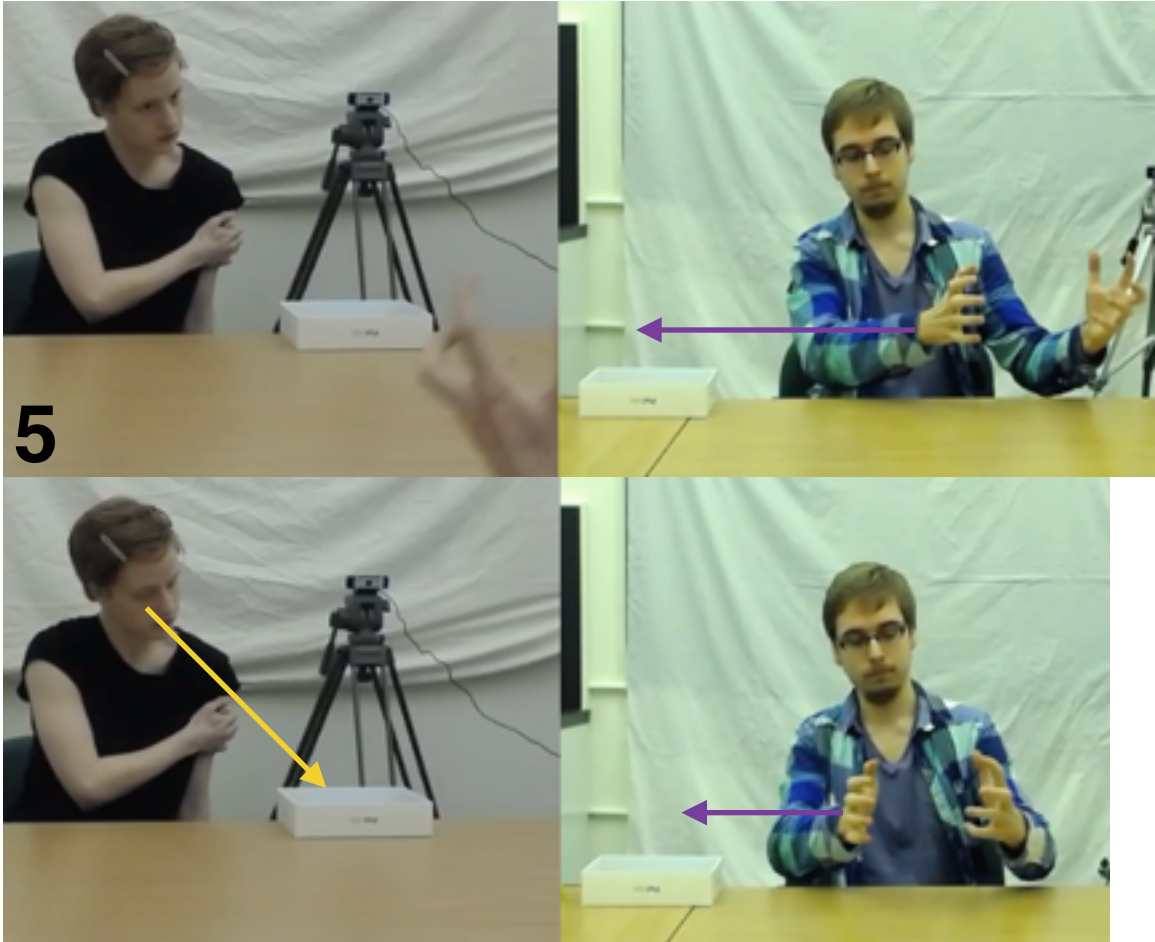
Image 3.7: Stills 5-6: Matcher Eye Gaze Eliciting Repair: Repair sequence has reformulation of path

If the sustained gaze is also combined with a furrowed brow- being a culturally shared index of confusion or misunderstanding- (as in Image 3.8 below), it could strengthen the miscommunication signal, therefore acting as a repair initiator. The Director (right) in Image 3.8 performs a sequential path-manner gesture. The first segment of the gesture is a partial pointed finger drawing the Slope manner (Still 1); then, the Director gestures the Roll manner, moving one hand over the other, but keeping the hands in place in front of her (i.e. not in a sloped trajectory) (Stills 2 and 3).
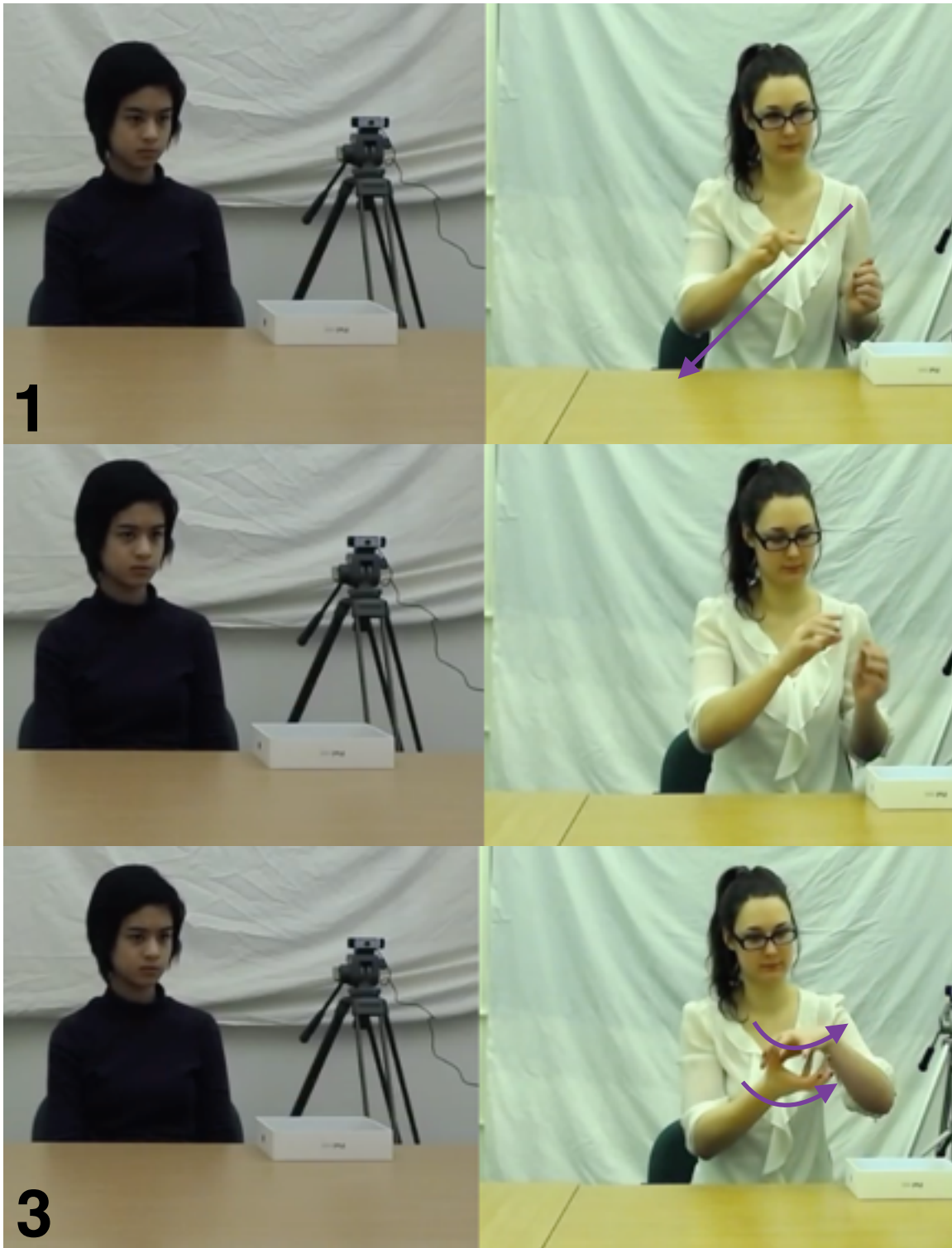
Image 3.8: Stills 1-3: Initial Gesture: Matcher (left) and Director (right) from Chain 4, Gen 2; Roll-Slope target

After the gesture completion, the Matcher (left) shifts her gaze to the screen, but instead of providing a guess, she furrows her brow to express confusion. The Director, who is looking at

the Matcher, can note this signal of non-understanding (Still 4), and respond accordingly. She begins a repair sequence (Still 5), re-gaining the gaze of the Matcher, with a modification of the path gesture: using her forearm to mark the Slope path, holding it in a diagonal line. The repair sequence is also combined with a slight smile from the Director, a possible sign of acknowledging the faulty communication- the smile is reciprocated in Still 6, as the Director continues the repair sequence. The second repair of this sequence has the Director performing manner and path simultaneously, as she "rolls" one hand over the other (mimicking the same Roll manner in the initial gesture), but adds the sloped trajectory to the manner gesture (Still 7). The repair gesture is successful, as the Matcher provides a correct response.

Image 3.8: Still 4: Repair Sequence Initiation: From Sequential to Simultaneous Gesture
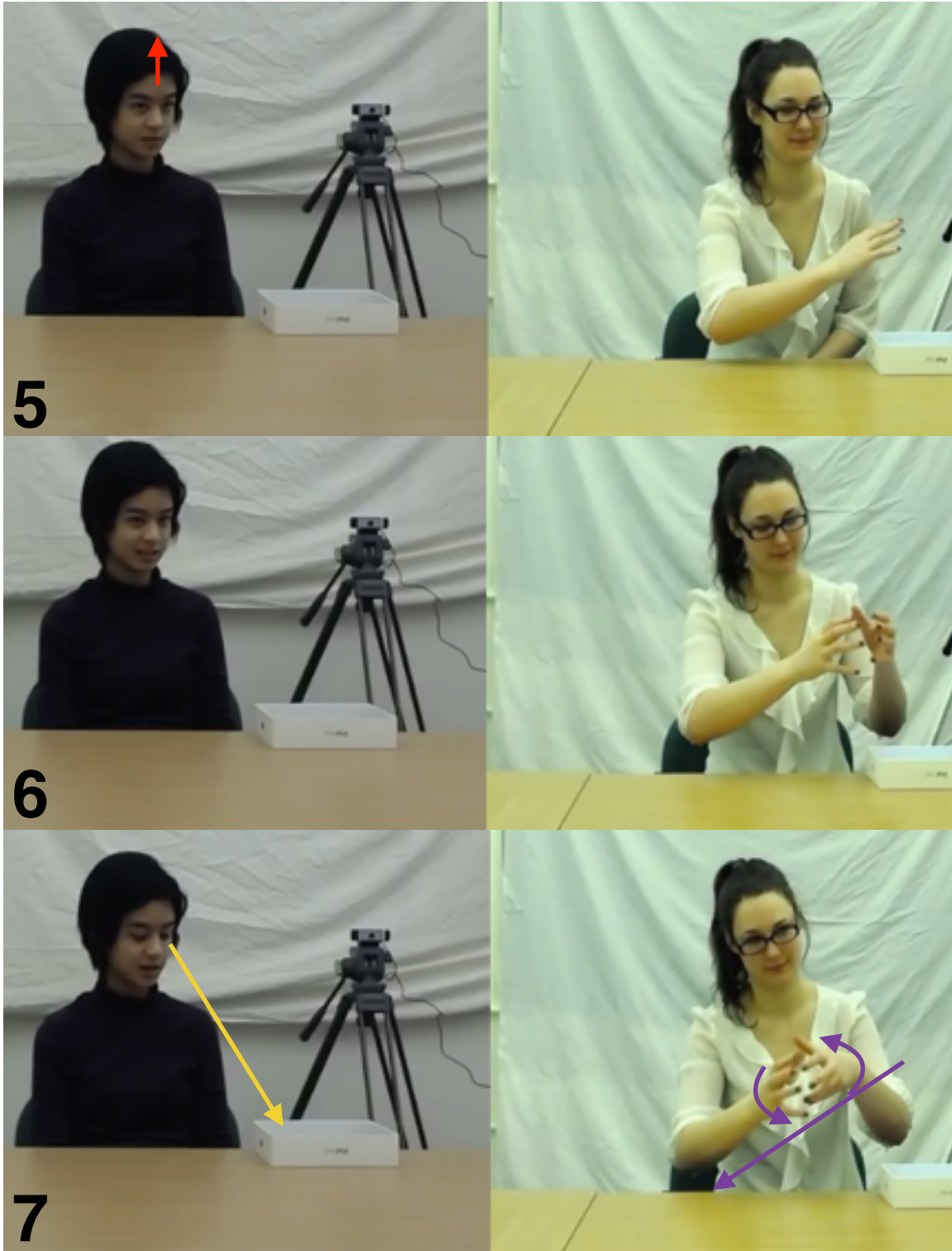
Image 3.8: Stills 5-7: Repair Sequence: From Sequential to Simultaneous Gesture

It appears the repair gesture in the above example was conforming to a structure that was

becoming increasingly popular within the chain. Initially, participant B (the Matcher in the above

example) had used sequential gestures in generation 1, but started to incorporate a sequential then simultaneous gesture structure early in generation 2. The observer from generation 1, now the Director in the example above, performed an initial gesture that conformed to the structure in generation 1. But, in the repair she switched on a simultaneous structure, possibly indicating a switch in structure preference, and convergence on one structure type. Repair allows the negotiation of structure, in which easily identifiable ones (more salient meanings) are preferred and reinforced. Repairs may be relied upon early in interactions to help establish a conventionalized system, and may be only be used for apparent miscommunication in later interaction.

Following the model proposed by Stivers and Rossano (2010), here mutual eye gaze between the Matcher and Director (or, at least the gesture space) extended the sequence. For example, if the matcher continued to look at the Director, the Director would continue to gesture. That is, a Matcher signals that the gesture is still relevant (to fully grasping the target meaning) by maintaining eye contact with the gesture or Director. If the Director also notes this continued gaze, they may repeat the gesture sequence or continue the current sequence (note that some sequences may be more suited for continuation, such as circle paths- we categorized repeats as a gesture that is taken up again after the hands have been at rest). The example below (Image 3.9), exhibits a continued gesture, though in a limited space, due to the gaze of the Matcher. While the Matcher (left) had initially been looking at the gesture, she diverts her gaze momentarily to look to the screen. In general, the Matcher's withdrawn eye gaze did not necessarily mean the Director would stop the gesture (this typically occurred either once the Matcher provided a guess, or, if the chain preferred it, after the first completed gesture). Given the nature of the

meaning space, gestures were sometimes performed in a continuous manner, and a Matcher's deterred gaze could be momentary, so Directors might chose to keep gesturing in case the Matcher returned their gaze to the gesture before guessing its meaning- as is down here. The Matcher quickly returns her gaze to the gesture space, though the Director has nearly completed the gesture (Still 1). The Director (right) gestures the manner and path, Spin-Slope, simultaneously moving one fisted hand (facing toward the Matcher) in a turning motion along the diagonally held forearm.
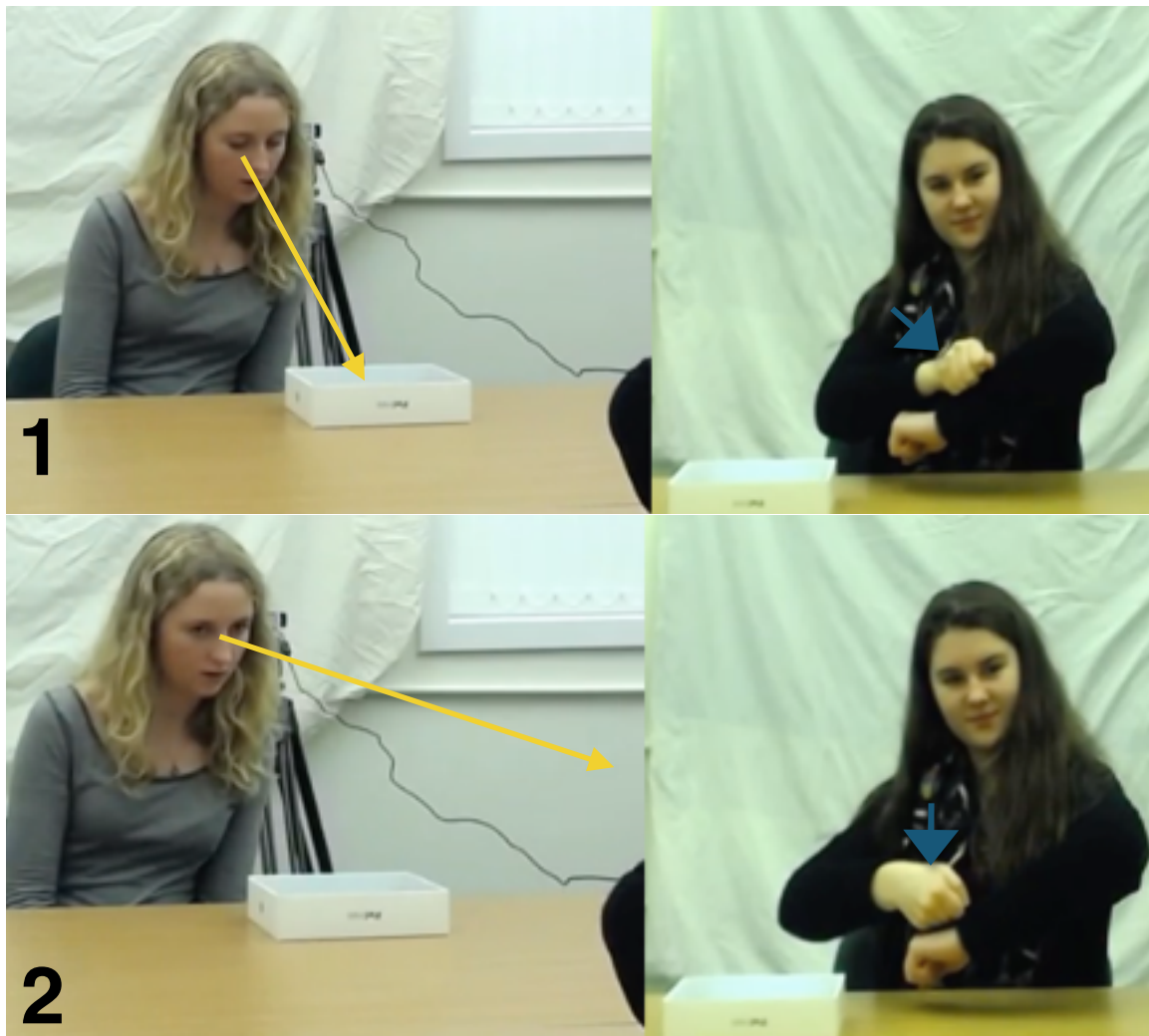
Image 3.9: Stills 1-2: Initial Sequence: Matcher (left) and Director (right) from Chain 1, Gen 3; Spin-Slope

As the Matcher returns her gaze to the gesture (Still 3), the Director has nearly reached the end of her diagonal forearm, yet seeing the Matcher is still attending to the gesture, she must provide more input. The Director shifts the axis of her fisted hand- gesturing the Spin manner- so that it faces downward, and giving the Matcher another perspective from which to glean the target meaning. While minimal, this clarification repair allows the Matcher to make a decision as to which image to select, and provides a correct response.
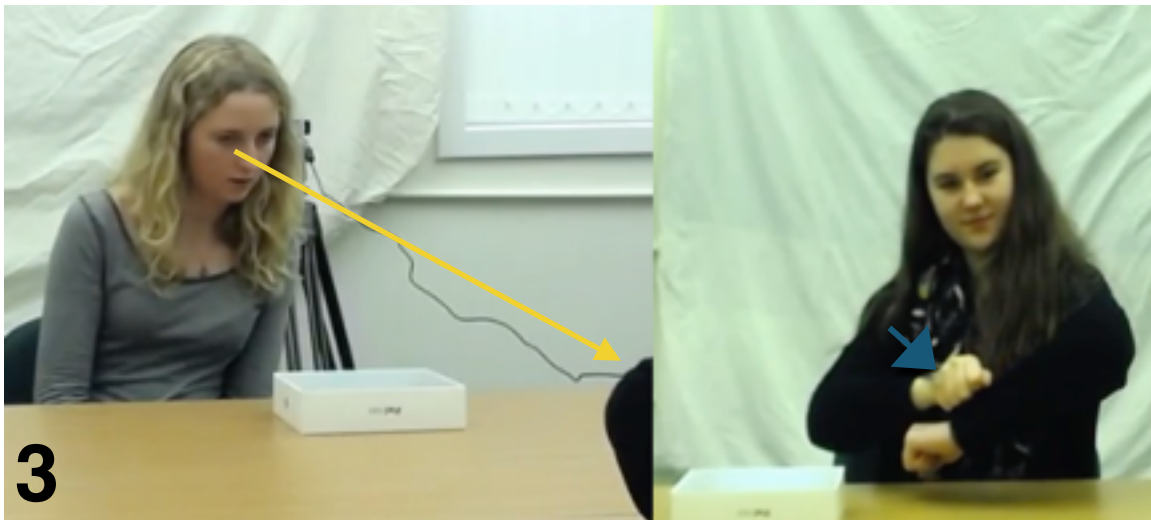
Image 3.9: Still 3: Repair Sequence: Director shifts the axis of the fisted hand

The Director took cues from the Matcher's eye gaze patterns to adjust her input, but preserved time by continuing to work in the gesture space remaining to her and making only the necessary repair of the Spin manner (the Slope gesture being quite salient as it is indexed twice, in the forearm and the hand trajectory). This examples shows not only a function of eye gaze in this experiment, but also the Director's awareness of the comprehensibility of their gestures.

Some Directors continued to gaze at the Matcher once the gesture had ended (as in generations that did not favor repeated gestures), even as Matchers looked to their screens. This seems to resemble the turn yielding strategy noted by Kendon (1967), wherein Directors allow

Matchers to take the floor and provide a guess. However, when Directors withdrew their gaze from the Matcher, it could indicate two functions. If the withdrawal occurred after the gesture was completed, but before the Matcher's guess, it could be a turn yielding strategy like that proposed by Goodwin (1980); only Matchers have the ability to make a next move in this interaction, so Directors can give up the floor by not maintaining their eye gaze. In contrast, if the withdrawal happens after the Matcher provides a guess, it would indicate a sequencing closing and a move to the next round. This preference was chain-determined, in which some chains exhibited lineage-specific eye gaze pattern (discussed above) which had them looking at the Matcher or their own gesture near the gesture completion and subsequent guess.

Eye gaze, an aspect of interaction that can only be co-opted for managing communication and negotiation in face-to-face, contingent interactions, is multi-functional in the case of emerging a communication system. As in natural conversation, eye gaze can demonstrate monitoring and attention-giving. In an emergent context, though, shifts in culturally-determined eye gaze patterns can signal trouble-sources and incite repair, which in turn can facilitate the conventionalization of the evolving communication system.

FUTURE DIRECTIONS

Not only was this study limited in terms of the communication that could be done between the dyads, but it also had complications during the experimentation. Some chains had to be eliminated from the results because they did not produce enough generations for analysis, technical issues with recording occurred, and the experiment room set-up was not ideal. The design also constrained the communication between the participants, who did not exhibit as much negotiation as pilot study participants had. Future studies would consider these limitations

93

and build upon the current study's goals to investigate the role of interaction in language evolution. Example designs, still using the gestural communication paradigm, include a card selection and trading task that would require negotiation between the dyad, as would using stimuli with confusable meanings (e.g. "hammer" and "hammering"). Each participant would have a number of cards they must obtain over the rounds of play, as with a "Go Fish" scenario, and must request the cards of their partner. However, given the confusable stimuli, they will be required to negotiate meanings and gestures to distinguish between similar phrases. This design could help inform the degree to which, and how, interactional features help evolve the communication system humans use to coordinate action.

CHAPTER 4

STUDY 2: GESTURAL COMMUNICATION OF NOUN-VERB PAIRS

The purpose of this experiment is to investigate the evolution of a communication system, along with other discourse features that accompany such a system, in a more openly interactive setting. The previous study (Chapter 3) examined these questions, albeit in a limited interactive environment. By expanding the interactive abilities and opportunities of the study's participants, we hope to gain more insights into specific aspects of language evolution, but in a setting that mirrors more natural language use. With a goal-based paradigm, we hope to approximate a context in which communication is required to benefit individuals, and in which negotiation of new concept meanings is required.

BACKGROUND

Research on evolving communication systems is varied in terms of methods, contexts, and modes of communication. The goal of some studies is to determine how communication emerges without a pre-designed system, while others look to how communication evolves over generations of users. The paradigms in past studies are varied, from using computer mediated tasks to pair-based guessing games. One feature that still remains less prevalent in these studies, however, is that of face-to-face interaction and its effects on language evolution and transmission.

Galantucci (2009), in a field called experimental semiotics, examines how novel communication systems are established and used to succeed in joint action tasks. The task takes place in a simulated world, played like a video game, wherein individuals must meet in the same room of the simulated world. This is accomplished by developing an effective communication

system using non-iconic graphic notations. The iconic notations used by participants were converted to non-iconic ones using a software that only recognized lines drawn on a certain axis/ horizon line. Greater success at this task relied on greater communicative power.. Consistently discriminating locations (which increased over rounds) within the simulated world, as indicated by the signs the pairs used, proved to be more indicative of success at the game.

A similar, yet more difficult, task has been developed by Scott-Phillips, Kirby, and Ritchie (2009) called the Embodied Communication Game. In this task, paired players had to develop a communication system from scratch in order to succeed at the task, which required them to move each individuals' icon to the same colored square on a computer mediated platform. The participants did not have access to an established communication system, nor to the randomized screen set up of their co-participant. The pairs that succeed relied on a default-based communication system through trial and error. One color, red, was the default choice for many pairs; when red was not an option, a participant would move their icon in a particular pattern to signal a different color option. While these experiments shed light on how communication systems emerge and become recognizable by their users, they do not consider the evolution of such a system as the users change over time.

The most notable method for tracking the evolution of a communication system has been iterated learning detailed in Kirby, Cornish, and Smith (2008). In models of iterated learning, participants learn an alien language in training, then use the language in a task that requires labeling images with the words they have just learned. The output from one participant is passed down to the next participant in the new training phase; the new participant also completes the labeling task and has their output transmitted to the next participants. One outcome of this model

96

is underspecification: using one label for a number of similar items. While easy to learn, there is much ambiguity in the system, limiting expressivity. For example, for meanings that differ in terms of color and movement but are the same for shape (e.g. "triangle"), color is underspecified as there is no index for it in the word produced. When the researchers introduced a hidden pressure to reduce underspecification in the participants' output (and subsequent training language), the outcome was a language that is systematic and compositional. That is, the language evolves from one-to-one idiosyncratic mappings to one that is easily learned as it becomes more compositionally structured (in which "the meaning of a given string could be inferred by the meaning of subparts of that string" (Cornish, Tamariz, & Kirby, 2009, p. 192)). For instance, color is marked with a prefix (ne = black, la = blue, ra = red), a triangle shape is indexed by the morpheme "ki," and movement is indicated with a final morpheme (ke = horizontal, pill = spiral). Over generations, language users simplify the idiosyncratic system to one that exhibits compositional structure. Compositionality allows for both a learnable and expressive language system.

Another application of iterated learning of communication systems is that of Fay, Garrod, Roberts and Swoboda (2010), though they extend the paradigm to include social collaboration. Using graphical sign system, participants, either in a pair or group, played a "Pictionary" type game in which they would guess the meaning drawn by another. Many meanings were quite abstract and could not be drawn with simple iconic representations (e.g. "Clint Eastwood" or "Parliament"). In an example of how a pair of participants represented "Parliament" over generations, it appears certain features common to each participants' first representation (curved lines likely resembling the seating in a Parliament building) become the anchor for more abstract

representations later in the generation. Through the iterated learning paradigm, the graphical representations become less iconic, while guessing accuracy shows the systems were successful in both the pairwise and community groupings. They conclude interaction was a crucial element to the creation of a new, shared communication system between individuals and communities of individuals.

Healey et al (2007) have sought to incorporate concurrent communication in their experiments, allowing for "mutual modifiability" from their participants. In a graphical drawing task, participants were given liberty to modify the each other's graphical input via a digital whiteboard tablet. The study was two-fold, investigating both interactions between different established sub-groups' communication systems as well as how the ability to modify inputs impacted how the drawings evolved over trials with either within or between subgroups. Individuals from the same sub-group produced more abstract drawings (deemed to be pre-compositional and more symbolic) than between sub-groups pairs, which produced more figurative (iconic) drawings. In support of the hypothesis for a conventionalized system emerging from interaction opportunities, they find "particular histories of interaction that accumulate within each group contribute to the emergence of localized, group specific, drawing conventions" (p. 297). Furthermore, interaction provided for the use of devices that supported coordination (agreement/disagreement), which often were done in abstract forms such as ticks, underlining, and arrows. These devices served editing functions that identified specific aspects of the pairs' drawings, creating alignment for meaningful elements of the representations.

In the second study by Healy and colleagues, manipulations to the extent interaction was allowed were employed for the same task as above. Participants in a blocked condition, in which

mutual modification of each others' drawings was not permitted, used more figurative drawings. The use of abstract drawings came about through opportunities for mutual modification interactions involving repair on co-participants' drawings rather than one's own output, as with the non-blocked condition. This finding is particularly intriguing, as it demonstrates clearly the importance of interaction to emerging communication systems between goal-directed individuals working together. Moreover, Healey and colleagues suggest the finding of "repair-driven co-ordination process" in communication could likely be observed in emerging gestural systems (and is widely demonstrated in vocal systems (Jefferson, 1983)).

Looking to gestural communication systems, Fay, Arbib, and Garrod (2013) consider the effectiveness of different modes of communication: non-linguistic vocalizations, gestures, and non-linguistic vocalizations plus gestures. To determine how effectively these systems convey motivated signs (i.e. their meanings can be determined through natural associations or resemblances) for emotions, actions, and objects, participants were put in Director-Matcher pairs randomly assigned to each communication condition. The pre-determined stimuli included easily confusable concepts, which Matchers had to determine from Directors' non-linguistic vocalizations and/or gesture. Gestural systems were more effective (successful) and efficient (time) than the non-linguistic vocalizations; furthermore, combining the two systems did not prove better than gesture alone. They conclude that gesture is a much more viable option for bootstrapping communication.

Previous studies have given us a wide range of resources to investigate the evolution of language and communication, but there still exists a lack in understanding of how interaction facilitates this evolution. The goal of the current study is to address this aspect of language use

and evolution, by seeking answers to the research questions below through a methodology grounded in previous studies, but broadened to consider the dynamic nature of interaction.

Research Questions

1. How do participants using an evolving communication system negotiate and conventionalize minimally contrastive concepts in the language?

2. How does the gestural communication evolve over generations? Does it evolve systematicity, compositionality, and expressivity?

3. How do features of interaction (eye gaze, facial gestures, etc) facilitate the evolution and transmission of the system? What are the functions for these features?

4. What is the role of repair in the negotiation of meaning and gestures in this task? How is it performed, recognized, and functionalized in this setting?

Hypotheses

Given findings from previous studies and our intuitions about language in interaction, we propose the following hypotheses to the research questions above:

1. Participants in early generations will use similar gestures for minimally contrastive concepts, but later generations will evolve distinct gesture-meanings for the concepts. Namely, they should develop a systematic and conventionalized gesture-meaning for the verb and noun categories that distinguish between them.

2. While the gestural communication system might start with and maintain pantomime (iconic) features, the necessity to disambiguate the minimally contrastive concepts will require a level of abstraction to the communicative gestures. That is, gestures that indicate action versus object may arise as compositional elements to the system.

3. As with the previous study, we hope to see interactional features transmitted throughout a chain of generations that are lineage-specific, particularly with regards to condition type.

We hypothesize that facial gestures will play a crucial role in detecting trouble sources as well as in the gesture-meanings themselves. Eye gaze, as in previous findings, will help direct attention to specific features of the gesture, as well as aid the monitoring of comprehension and success in the negotiation.

4. Given the more open interactive task here, we hypothesize that repair will be more prevalent than in previous studies. Repair will necessarily be done via gesture (likely accompanied by eye gaze and/or facial gestures)- this may come in the form of repeats or clarification of gestures. Repair will play a pivotal role in negotiating meanings.

METHODOLOGY

In an experimental setting, participants engaged in a interactive communication game aimed at producing an evolving communication system through negotiation over simulated generations. Participant dyads were required to interact with one another during a card ("word") selection and trading task, akin to "Go Fish." However, their communication was limited to silent gestures (namely with the hands/arms, but may also include the face), similar to "Charades."

Participants

All participants were non-signing native English speakers, university-aged, and right-handed. They were recruited through flyers, undergraduate classrooms, and other university-approved solicitation methods. Three chains of 4 dyadic "generations" were collected and recorded per condition (described below) for this experiment, for a total of 36 participants. Participants were reimbursed for their time.

Stimuli

The stimuli for this experiment were two sets of similar-meaning cards. One set was comprised of verb meanings, while the other was comprised of noun meanings. However, the verb-noun pairs had minimally contrastive features when gestured with speech (determined from spontaneous co-speech gestures as well as elicited gestures from a pilot study). Also, Clark & Clark (1979) provided numerous examples of nouns that have been transformed into verbs in natural language use, and while not all the examples conform to the standards for this experiment, we did use many examples in the current stimulus set. Specifically, the instrument category has greatly influenced the stimuli. A sample verb-noun distinction might be "hammering" and "a hammer", respectively.

The set of stimuli included the noun/verb distinctions shown in Table 4.1 below, categorized according to Clark & Clark (1979). The first and second columns comprise the target stimulus set, 32 pairs totaling 64 individual tokens. The third column items were distractor tokens that contributed to task difficulty and masked the task objective (i.e. some did not conform to the grammatical forms of the target noun-verb pairs). Distractors were from similar semantic categories as the target words. The verb cards were grammatically represented in the progressive form ("-ing" suffix), while the noun cards adopted the indefinite article "a" or a null article (as with "snow").

Instruments	Elements, Actions, Misc.	Distractors*
(a/to) Bicycle	(to) Snow	A Comb
(a/to) Nail	(a/to) Guard	Tools
(a/to) Hammer	(a/to) Dream	A Playground
(a/to) Drum	(a/to) Picture	A Camera
(a/to) Zip(per)	(a/to) Cook	A Kitchen
(a/to) Shield	(a/to) Garden	Cleaning
(an/to) Iron	(a/to) Spray	A Referee
(a/to) Skate	(to) Milk	A Gym
(a/to) Phone	(a/to) Peel	Construction
(a/to lift) Weight	(a/to) Slide	A Cow
(a/to) Rake	(a/to) Swing	Combing
(a/to) Ski		Cheese
(a/to) Snowboard		A Paddle
(a/to) Brush		Paddling
(a/to) Grate(r)		Sleeping
(a/to) Shovel		
(a/to) Saw		
(a/to) Photograph		
(a/to) Vacuum		
(a/to) Whip		
(a/to) Whistle		

Table 4.1: Noun-Verb Stimuli

Procedure

The participants were video-recorded, as were the card selections of each participant in each round. The experimental set-up was designed to facilitate face-to-face interaction, with minimal obstruction of the hands, upper body, or face.

Dyads, using only silent gestures, played an interactive card selection task for four rounds (per generation). Each participant in the dyad had a target card set (n=8) which they attempted to convey by the end timed rounds. They also had a set of cards "in hand" (n=10, 8 target matches,

and two similar-semantic distractors) which could only be viewed by the holder at any point. For each round, the target set of noun-verb pairs were grouped for similar semantics (e.g. "shovel" and "rake" cards were in the same round), but were distributed between the participants randomly (for noun-verb pairs, as well as for the number of nouns or verbs in each participants' set).

To acquire the target card set, participants took alternating turns requesting one card at a time with silent gesture (ordered from 1-8 for each participant, in which they could only view one at a time and could not move ahead) from their partner's cards-in-hand. They did not need to construct gestures for the actual request (they were told the goal before the start of the experiment).

Participants were placed in one of two conditions: only distinguished by the ability of the Director to perform a "do-over" turn immediately following an incorrect guess. In the "Standard" condition, each participant had one opportunity per turn to make a request for the target card. Once the gesture was performed (note that the gesture could be performed or changed any number of times before the Matcher provides a guess), the Matcher provided a card, shown face-up for all participants present. If the card was a correct match, the participants placed it to the side (a "matched matrix"); however, if the card was incorrect, the Matcher returned the card to their in-hand board, and the Director placed the target card to the bottom of their set. The Director would have another opportunity to gesture the target card later in the round. The "Do-Over" condition proceeded in the same manner, expect in the instance of an incorrect guess from the Matcher. In this instance, the Director had an immediate turn for a "do-over" in which they

could once again gesture the target meaning. This condition was meant to drive immediate repair and reflect how repair is performed in natural conversation.

An Observer was also be present in the experiment room, in a non-interacting role. Instead, the Observer was required to observe the gestures and interaction between the dyad, knowing they would be the next to participate with one of the members they observed. Observers also had sight of the cards traded between the dyads, but not the target sets. When the interacting dyad's rounds finished, the Observer joined one of the previous members (the most "novice") to form a new dyad, and another Observer entered.

Coding and Measures

The video-recorded data were coded for handshape, marker, and repair performance and initiation. Handshape was coded broadly, using the categories Action, Handled, Embodied, Shape, Mov, and Loc. "Action" referred to the performance of the target word, while "Handled" was used exclusively for noun targets as a means to measure the difference between object-variants (also called perceptual variants) and action-variants (Ortega, Sumer, & Ozyurek, 2012). "Embodied" hand-shapes were those in which the participant used part or all of their body as the object. "Shape" referred to the use of the hands to either trace or represent the shape of the target. "Mov" referred to the use of the hands to indicate the movement of an object, usually in opposition to the participant performing the action. And, "Loc" was used to indicate a reference a location of the target (e.g. on the feet). Repairs were classified as either repetitions, clarifications, or reformulations. While these classifications are derived from conversation analysis (Jefferson, 1972), the application of the schema to gesture-based work does involve some interpretation on the part of the researcher (for example, in determining how precise the repetition of a gesture

105

must be before it is considered a clarification). Gesture length (timing in seconds) and accuracy was also measured and coded.

Post-test questionnaires were administered to gather first-hand accounts of how successful the participants believed the communication to be, how they changed strategies, if they recognized misunderstandings and repairs, and the extent to which they could distinguish their partner's gestures for the noun-verb categories. These answers were collected through both Likert scale measures as well as free written responses (see Appendix C).

RESULTS

Here we will detail the results of the study, namely *what* gestures the participants produced to communicate, disambiguate, and repair meanings. Furthermore, data on timing and accuracy will be provided for chains and generations. Finally, statistical analyses for variables will be determined.

Gestured Meanings: Markers to Disambiguate Noun-Verb Pairs

We hypothesized negotiation of meanings and gestures will be required to distinguish between verb-noun pairs, and that the participants will need to conceive of and conventionalize a set of gestures that either differentiate the quality of verb and noun, or devise highly individualized gestures for each item in the set. The latter solution was unlikely to occur due to its inefficiency regarding memory and learning. To innovate and faithfully transmit 64 individual gestures would prove time-consuming and difficult. Therefore, the more likely solution, in which a conventionalized system for distinguishing between nouns and verbs is developed, was the preferred solution for the participants of both conditions in this study.

Given our intuitions about the differences between nouns and verbs in gesture- namely co-speech gesture- and the grammar of many spoken languages, we hypothesized specific differences would be found in how participants distinguish between the noun-verb pairings. In many languages, verbs tend to receive more morphological inflections to denote person, plurality, and tense, however, verbs received seven times less marking than nouns (416 marked nouns: 70 marked verbs). As is seen in some signed languages (American Sign Language, for example), repetition was the most frequent manner in which to mark a verb. Image 4.1 below is a verb-marking repetition sequence in which "Snowing" is communicated by two hands with articulated fingers moving downward (stills 1-2), followed by a gesture iconic of "being cold" (stills 3-5). The action sequence is then repeated with fingers articulating snow falling downward (still 6). The Do-Over condition only used this marker for verbs (n=19), while it remained the most common in the Standard condition (n=35). However, unlike ASL, the repetition within the gesture was not systematic, specifically with respect to its noun counterpart. That is, noun gestures were also repetitive, while verb gestures were more exaggerated in their repetition (some participants even included facial gestures to accompany prolonged verb gestures). A second, and infrequent (n=3), verb marking strategy in the Standard condition was the rolling of the fingers to mimic a repeated or continuous occurrence. This marker could be a result of the progressive form in which the verbs were presented. The "-ing" suffix, informally called "the progressive" form, is then displayed in an iconic gesture of ongoing action.
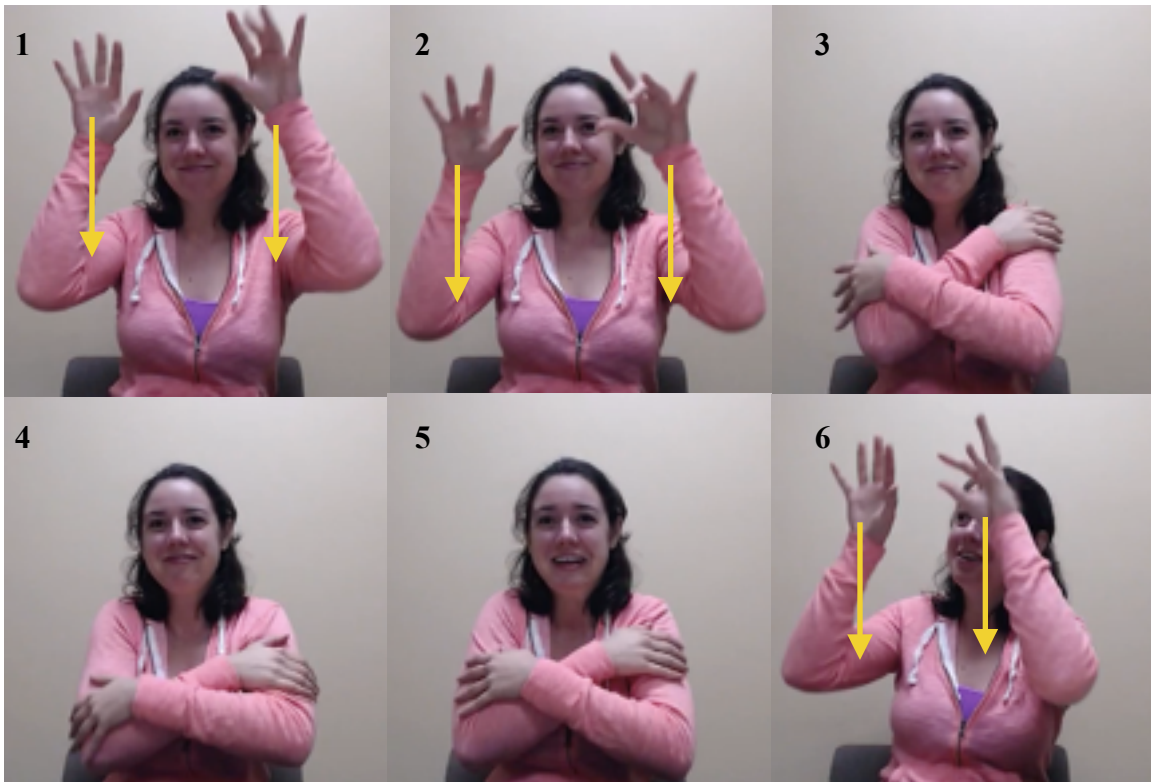
Image 4.1: Repetition Verb-Marker: "Snowing"

A final verb marker was the "hand-fan" (n=13), or a communication of the entire event as displayed through one hand fanning over the action of the other hand in a circular motion. The "fan" covers the gesture space, as though the Matcher should consider the "whole" of the space, not just an singular item within it, as is done in Image 4.2 (stills 4-5) after "Drumming" has been gestured through action (stills 1-3). This marking system was only present in one Standard chain, but lost favor as a more recognizable noun-marking system came into play. Verbs may have been, in general, un-marked due to the nature of hand-shapes the participants adopted: action-variants. That is, the semantic information was conveyed through the handshape itself, and marking was not deemed necessary for meaning detection.

Image 4.2: "Hand-fan" Verb-Marker: "Drumming" is gestured by performing the action and then swirling the hand in a circular motion over the gesture space.

Nouns, on the other hand, received much more variety and frequency of marking in both conditions. We hypothesized that a point to the gestured noun would distinguish it as the object, as if pointing to a real object in the imagined context. For example, following a hammering action gesture (Image 4.3, stills 1-2), the Director points to the target noun, "A Nail," which is being "held" in the non-action performing hand. Chain 3 in the Do-Over condition and Chain 2 in the Standard condition (i.e. one chain in each condition) primarily relied upon this strategy to mark a noun, and many were in fact aware of the use of this marker. Furthermore, a distinct word order developed in noun gesture production, in which an object-point was inserted into the action sequence typical of the verb gesture. Participants noted the need to wait to determine if their partner would point to an object after having done an action gesture or if they would continue with the action.
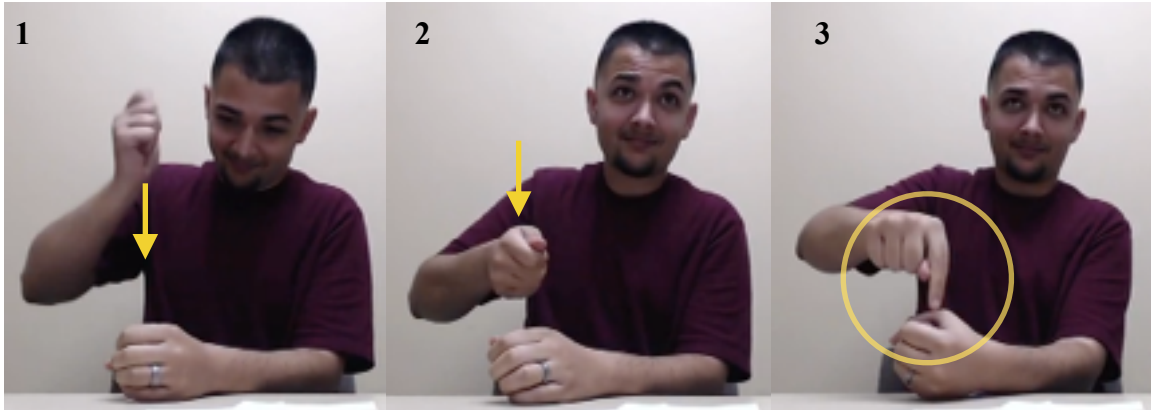
Image 4.3: O-Point Noun-Marker: "A Nail"

While the object-pointing strategy was used in many chains, other chains adopted different   dominant strategies for noun-marking. Similar to the object-point, two chains (one from each condition), co-opted the index finger, the number "1," to represent "a" as displayed on 30 of 32 noun cards. As in Image 4.4, an index finger is held up first to mark the noun-ness of the target, "An Iron." Then, the subsequent handled gesture has an imagined iron move across the table. In fact, one chain in the Do-Over condition so reliably used the index finger to mark singularity (or noun-ness) that they generalized it to the non-count nouns "snow" and "milk" by the final generation within a chain. The recognizability of this marking form is a likely reason for it to overtake another noun-marking form mid-chain, and then remain the dominant noun marker.
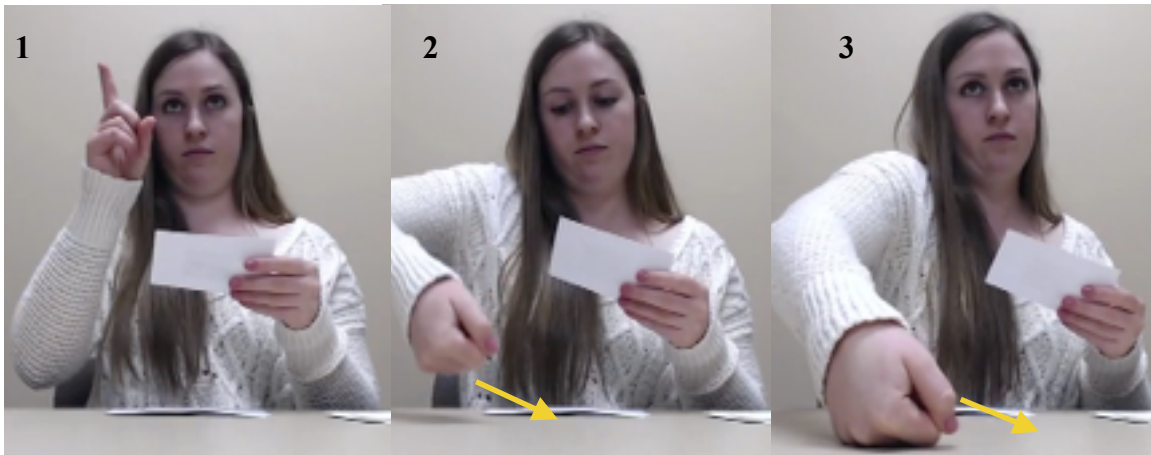
Image 4.4: Index Finger Noun-Marker: "An Iron"

A final form of noun marking, object emphasis, took on many forms, including an emphasis on handheld-ness, embodiment, and shape. Together, the object emphasis marker was used 91 and 81 times in the Do-Over and Standard conditions, respectively. The handled variant of an object emphasizing marker involved a participant performing the action variant and then emphasizing the object used to perform that action, usually by shaking the handling hand. In Image 4.5, a participant first gestures "A Grater" with one hand holding the imagined object and the other enacting the related verb (stills 1-3). After this sequence, the hand "holding" the object remains in the gesture space and shakes back and forth to highlight the imagined in-hand object.

Image 4.5: Object-Emphasis Handled Noun-Marker: "A Grater"

Another means to highlight object-ness was through embodiment. Verbs were rarely performed through embodiment, or "body-as-object," but nouns were more likely to be embodied if they could be instrumentally performed, as in Image 4.6. Here, the participant holds her arm and hand in a straight line (still 1) as "The Saw" target (compared to, for example, the handled saw). She then moves her arm back and forth in a sawing motion. As verbs are not often embodied, this body-as-object handshape emphasizes the object rather than the action being performed.
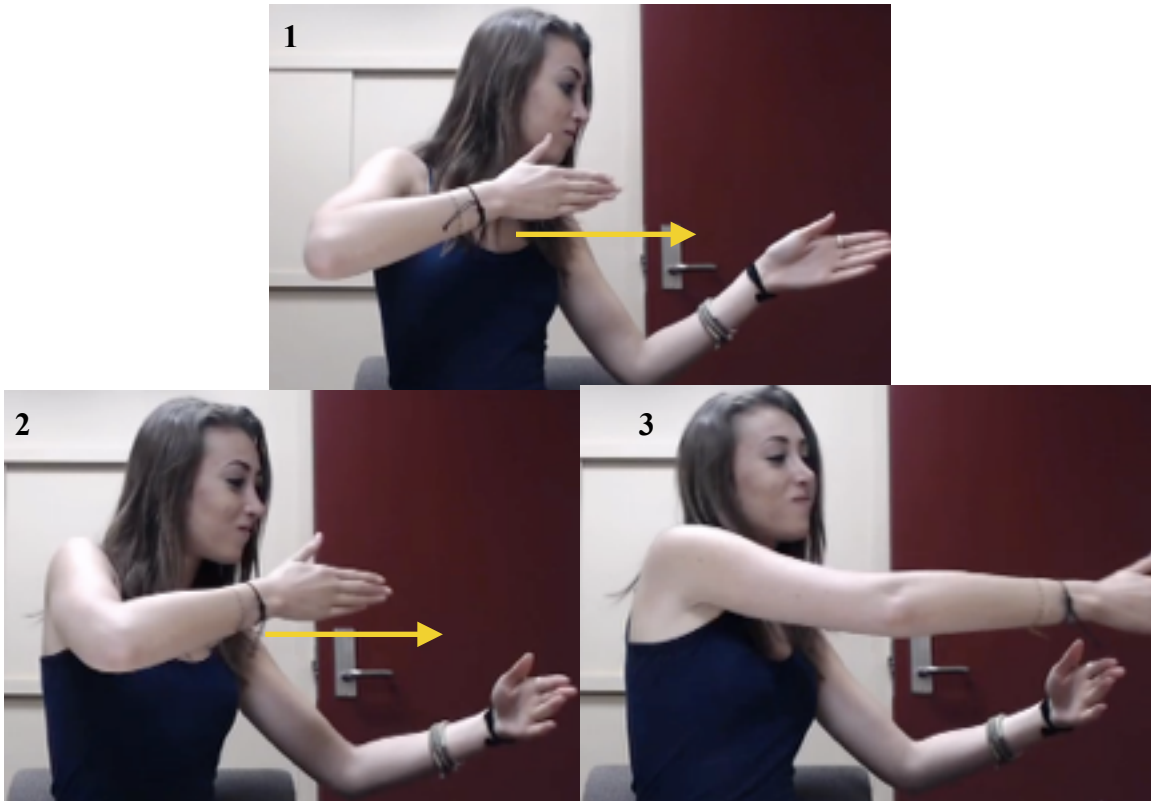
Image 4.6: Object-Emphasis Embodied Noun-Marker: "A Saw"

A third emphasized feature was shape, which co-occurs often with the "Shape" handshape to be

discussed below, but emphasis was not necessarily a characteristic of the handshape "Shape".

However, the shape displayed in the object marker often did not resemble the object's actual

shape, but rather indexed "item-ness" through a generic box shape. This particular object

emphasis was not common (n=2), but did successfully convey noun-ness. Object emphases were

often exaggerated with co-occurring facial gestures, such as raised eyebrows, which were a

redundant highlighting feature but a nonetheless salient one for Matchers to detect.

Patterns in noun-marker use varied across individual chains, though Do-Over chains

appear to align to one strategy earlier than Standard chains. Figure 4.1 below illustrates which

markers were used in each chain and their frequency.
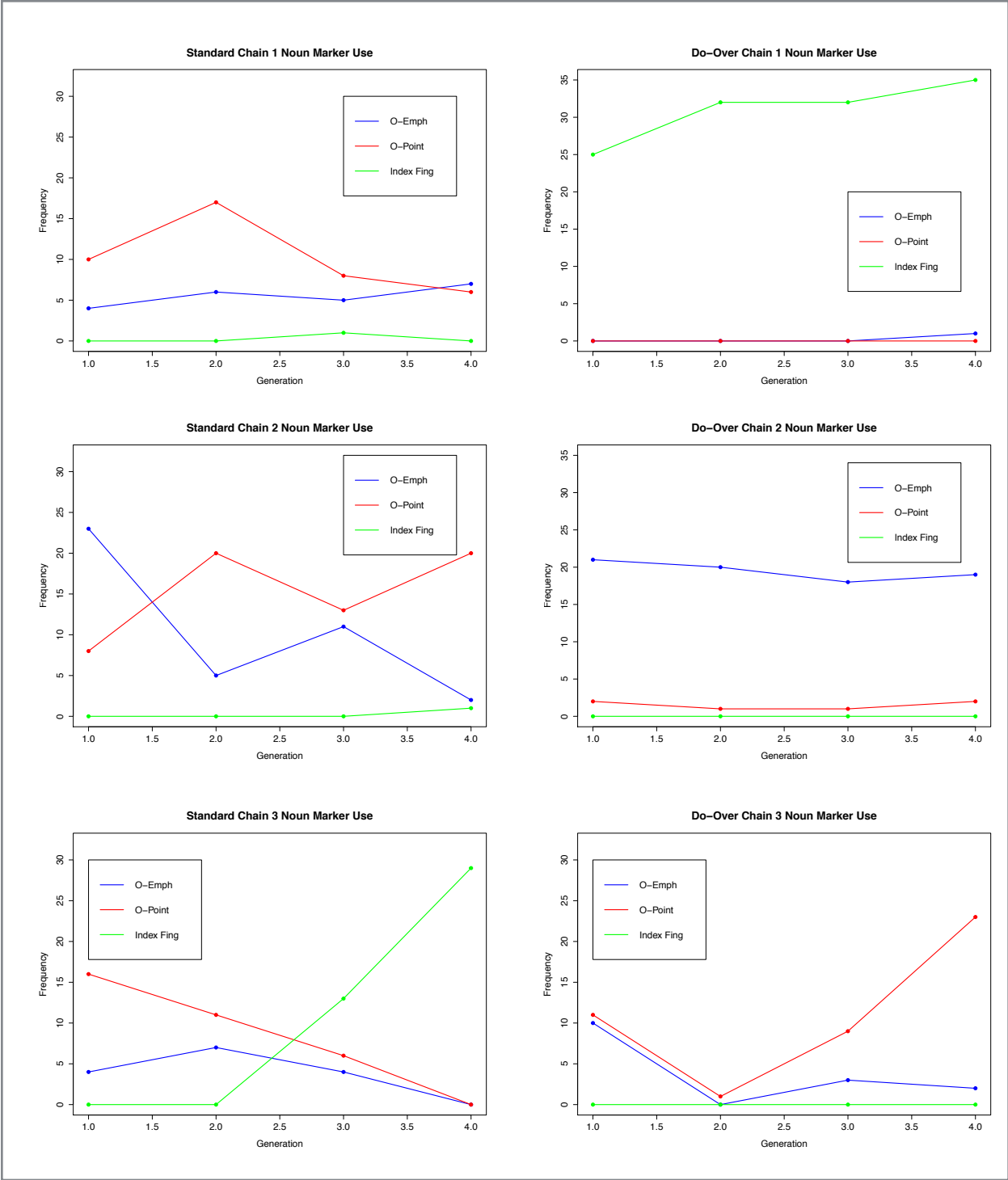
Figure 4.1: Noun-Marker by Chain: Do-Over Chains 1-3 are on the left, Standard Chains 1-3 are on the right

Do-Over chains exhibited less variability in marker use, namely in Chains 1 and 2, while

Chain 3 showed more alignment to a dominant marker in Generation 3. All Standard chains had

non-dominant use of markers initially, wherein Chains 2 and 3 displayed a switch in the marking strategy that eventually overtook the other form in use, occurring in generations 2 and 3, respectively. A generalized logistic regression was run to determine if generation predicted the presence of individual marker strategies (against no marker use) for nouns in a given chain.

Standard	Chain 1					
	O-EMPH			**O-POINT**		
	z value	*Pr(>|z|)*	*Probability*	*z value*	*Pr(>|z|)*	*Probability*
Gen 1	-8.009	1.15e-15 ***	0.16759777	-1.989	0.0467 *	0.3103448
Gen 2	-0.612	0.54027	0.14173228	2.563	0.0104 *	0.6818182
Gen 3	-1.324	0.18539	0.10891089	1.835	0.0666	0.6666667
Gen 4	-2.982	0.00287 **	0.05517241	-1.476	0.1399	0.1428571

Standard	Chain 2					
	O-EMPH			**O-POINT**		
	z value	*Pr(>|z|)*	*Probability*	*z value*	*Pr(>|z|)*	*Probability*
Gen 1	0.684	0.4937	0.5588235	-1.436	0.1510	0.3478261
Gen 2	-1.060	0.2893	0.3846154	2.551	0.0107 *	0.7142857
Gen 3	-1.881 .	0.0600	0.2000000	1.776	0.0758 .	0.6190476
Gen 4	-2.527	0.0115 *	0.1333333	1.779	0.0753 .	0.5937500

Standard	Chain 3					
	INDEX FING			**O-POINT**		
	z value	*Pr(>|z|)*	*Probability*	*z value*	*Pr(>|z|)*	*Probability*
Gen 1	-0.01	0.992	3.181005E-09	-2.073	0.0382 *	3.409091E-01
Gen 2	0.00	1.000	3.181005E-09	0.228	0.8198	3.666667E-01
Gen 3	0.01	0.992	5.909091E-01	0.412	0.6801	4.000000E-01
Gen 4	0.01	0.992	7.941176E-01	-0.011	0.9910	2.350463E-08

Table 4.2: Standard Condition Chains: Individual Marker Strategies were compared against no marker use for noun targets in individual chains.

Table 4.2 (above) shows the differences in a given marker strategy over generations, as well as the predicted probabilities of each marker (compared with no marker use at all for noun targets). Standard Chain 1 had a significant change (increase, in this case) in O-Point marker use from generation 1 to generation 2, as did Chain 2. Given the nature of marker use in Chain 3, in which the Index Finger marker had no presence in generations 1 and 2 to becoming the only form used (and ubiquitously used for nouns), the model does not properly reflect the effects. Nonetheless, the Index Finger marker does have a dramatic increase in use starting in generation 3, as the other two forms steadily decline.

Table 4.3 reflects the same model data as Table 2 above, but for the Do-Over condition.

Do-Over	Chain 1			Chain 2		
	INDEX FING			O-EMPH		
	z value	Pr(>|z|)	Probability	z value	Pr(>|z|)	Probability
Gen 1	1.564	0.11780	0.6250000	0.156	0.876	0.5121951
Gen 2	2.408	0.01602 *	0.8823529	0.806	0.420	0.6060606
Gen 3	2.790	0.00527 **	0.9393939	-0.105	0.916	0.5000000
Gen 4	0.010	0.99164	1.0000000	-0.560	0.576	0.4500000

Do-Over	Chain 3					
	O-EMPH			O-POINT		
	z value	Pr(>|z|)	Probability	z value	Pr(>|z|)	Probability
Gen 1	-2.190	0.0285 *	2.857143E-01	-2.391	0.0168 *	0.25925926
Gen 2	-0.010	0.9917	3.181005E-09	-2.293	0.0218 *	0.02702703
Gen 3	-1.975	0.0483 *	8.571429E-02	-0.998	0.3181	0.15789474
Gen 4	-1.005	0.3147	1.428571E-01	2.920	0.0035 **	0.64705882

Table 4.3: Do-Over Condition Chains: Individual Marker Strategies were compared against no marker use for noun targets in individual chains.

Chains 1 and 2 had only one marker present, and those results have been included in a shared table section. The Index Finger marker in Chain 1 demonstrated a significant difference (increase) in use over generations, until it reached conventionalization (used for all noun targets) in generation 4. In contrast, O-Emph in Chain 2 was used to a similar extent in all generations, not increasing or decreasing in frequency significantly over generations. Chain 3 began with the split use of O-Emph and O-Point marking, however, over generations, O-Point saw a more significant change in use, wherein it became the dominate marking form by generation 4. For a comparison of differences between more commonly used marker strategies with a chain, see Appendix D.

Predicted probabilities were determined from a generalized logistic regression model, with the same parameters as the above data. In Table 4.4, we considered the predicted probability of any noun marker use by chain and by condition over generations. Do-Over chains are more likely to have increased probability of noun marker strategies from generation 1 to generation 4. Chain 2, however, appears to maintain a relatively consistent use (about 50%) of one noun marker, which may be linked to the individual nouns themselves. Nouns that are more difficult to distinguish gesturally from their verb counterpart may receive a noun marker, but nouns more distinguishable from their associated verb may not receive the same marker. That is, this chain did not generalize the marker strategy to all noun-types. In contrast, only one Standard chain resulted in above 50% use of noun markers by the final generation; this same chain (3) exhibited a dramatic shift in marker strategy, from a mixed use of O-Point and O-Emph (in Generations 1 and 2) to an exclusive use of Index Finger by generation 4. Other Standard chains had a varied use of O-Point and O-Emph throughout the lineage, meaning theses chains did not

conventionalize to one marking system. Collapsing all chains in a condition, we see the Standard condition was nearly 60% likely to use a noun marker by generation 4, while the Do-Over condition was 70% likely to do the same (for comparison probabilities, see Appendix E).

	Standard 1	Standard 2	Standard 3	Overall Standard	Do-Over 1	Do-Over 2	Do-Over 3	Overall Do-over
Gen 1	0.4117647	0.5121951	0.4081633	0.4920635	0.6250000	0.5348837	0.48717949	0.5491803
Gen 2	0.7666667	0.3970588	0.4864865	0.6600000	0.8823529	0.6176471	0.02702703	0.4952381
Gen 3	0.8125000	0.4285714	0.7352941	0.7647059	0.9393939	0.5142857	0.23809524	0.5363636
Gen 4	0.3142857	0.3766234	0.8000000	0.5963303	1.0000000	0.4883721	0.67567568	0.7043478

Table 4.4: Predicted Probabilities of Marker Strategy Use in Noun-Type Targets: Over Generations by Chains

Returning to Tables 4.2 and 4.3, predicted probabilities of individual marker strategies (compared to no marker use) revealed finding that support those in Figure 1. For the Standard condition, Chain 1 markers became less predicted for use in Generation 4, while Chains 2 and 3 were approximately 60% likely to use O-Point and 80% likely to use Index Finger markers, respectively, by generation 4. In the Do-Over condition, Chain 1 was 100% likely to use the Index Finger marker by generation 4- indicating its reliability in marking noun-ness. Chain 2 was 50% likely to use the O-Emph marker by the end of the chain, suggesting its less aggressive application to all noun targets.

Overall, the Standard condition marked targets more than the Do-Over condition. The Standard condition participants marked more nouns than were left unmarked, and while the primary overall strategy was pointing to the imagined object, each chain exhibited different dominant strategies.

Gestured Meanings: Noun Handshape

As we hypothesized, the semantic category to which a noun belonged could influence how its gestural structure is manifested. Instrument nouns (23 of 32 noun targets, such as "A Hammer" or "A Shovel") are most often gestured as "Handled," that is the participant holds the imagined object in their hand as they perform the action related to the instrument. This is an action-variant (Ortega et al, 2012), by which an object is represented by the action associated with it. A notable exception is "A Phone," which is most frequently embodied as a hand in the shape of a phone. Overall, nouns were nearly evenly divided between "Action" and "Handled" hand-shapes, which could be attributed to the semantic category. "Shape" and "Embodied" hand-shapes were the third and forth, respectively, most frequent for both conditions. Overall, there was a trend to use action-variants ("Handled") over object/perceptual variants ("Shape" or "Embodied") for noun targets.

More varied noun handshape depended on semantic category, while verbs, regardless of category, were nearly exclusively gestured by an "Action" in the Standard and Do-Over conditions (373 of 463 verbs, and 369 of 433 verbs, respectively). "Shape" and "Mov" were the second and third, respectively, most frequently used handshapes for verbs in both conditions. The "Handled" handshape was least frequently used.

Repair

All three repair strategies were present in the two conditions, though to differing degrees (Figure 4.2). Clarifications were the most frequent strategy in either condition (Standard=52, Do-Over=63). A clarification repair required the Director to highlight or emphasize an element of the initial gesture, typically done through the shaking of a hand or emphasizing a marker;

119

clarification repairs are moderate modifications on the initial (trouble-source) gesture. Reformulations were more frequent in the Standard condition (n=23) than the Do-Over condition (n=9), which may be related to the inability to perform a repair after an incorrect guess, thereby requiring Directors to re-do gestures without knowing what the trouble source or misinterpretation is. Reformulations are gestures that are complete modifications of the initial gesture, often involving a change in handshape. Repetition was similarly used in both conditions (Standard=21, Do-Over=20), and required no modification to the initial gesture. Repetitions could be full or partial repeats of the initial gesture, but crucially did not involve emphasis in the repeated gesture.
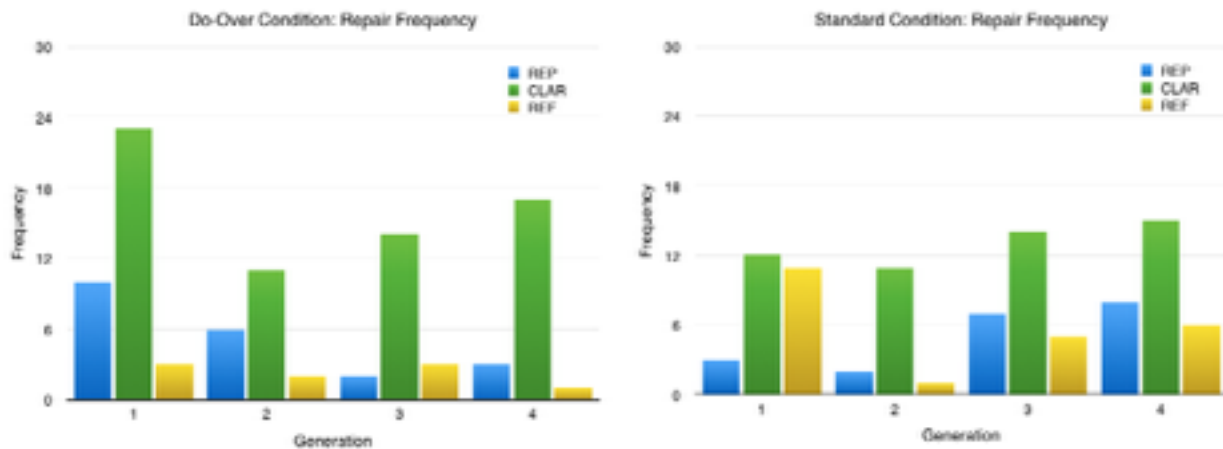
Figure 4.2: Repair Use by Generation: Do-Over condition on the left, Standard condition on the right; repair strategies (from left to right) are Repetition, Clarification, and Reformulation

An example of a clarification repair is detailed in Image 4.7 below in which a Director emphasizes the target by shaking the imagined handled object. Here, "A Weight" is gestured with a "Handled" handshape and curled up from a horizontal to a vertical position (stills 1-2). The hand "holding" the imagined weight rotates to face the Matcher, as a means to indicate they should guess the object-in-hand (still 3). An incorrect guess and the Director rejection (still 4) allows the Director (from a Do-Over condition) to highlight the noun-ness of the target word by

120

shaking her hand (still 5) to indicate the object in her hand is the intended referent. This clarification results in a correct guess.
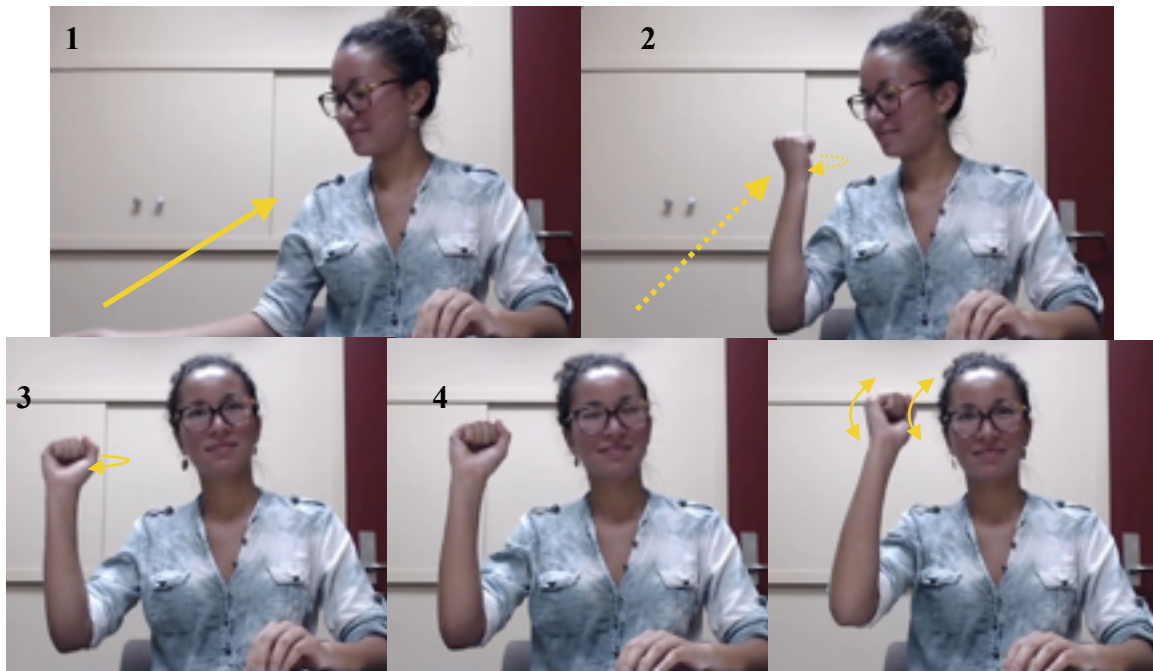
Image 4.7: Clarification Repair: "A Weight" is gestured and then displayed, and is clarified by an emphatic shake.

A reformulation, in contrast, repairs the initial gesture in such a way that it becomes a wholly new gesture, as seen in Image 4.8. The Director's initial gesture for "A Shield" is composed of a trace of the shape of a shield (stills 1-3) and the handling of the drawn shape with both hands in front of her body (stills 4-5). When this gesture does not efficiently communicate the target meaning, the Director reforms the gesture through an action of shielding or protecting oneself (still 6). This reformulated gesture is effective in its communicative intent and, in fact, remains a gestural strategy for "A Shield" in the subsequent generations.

121
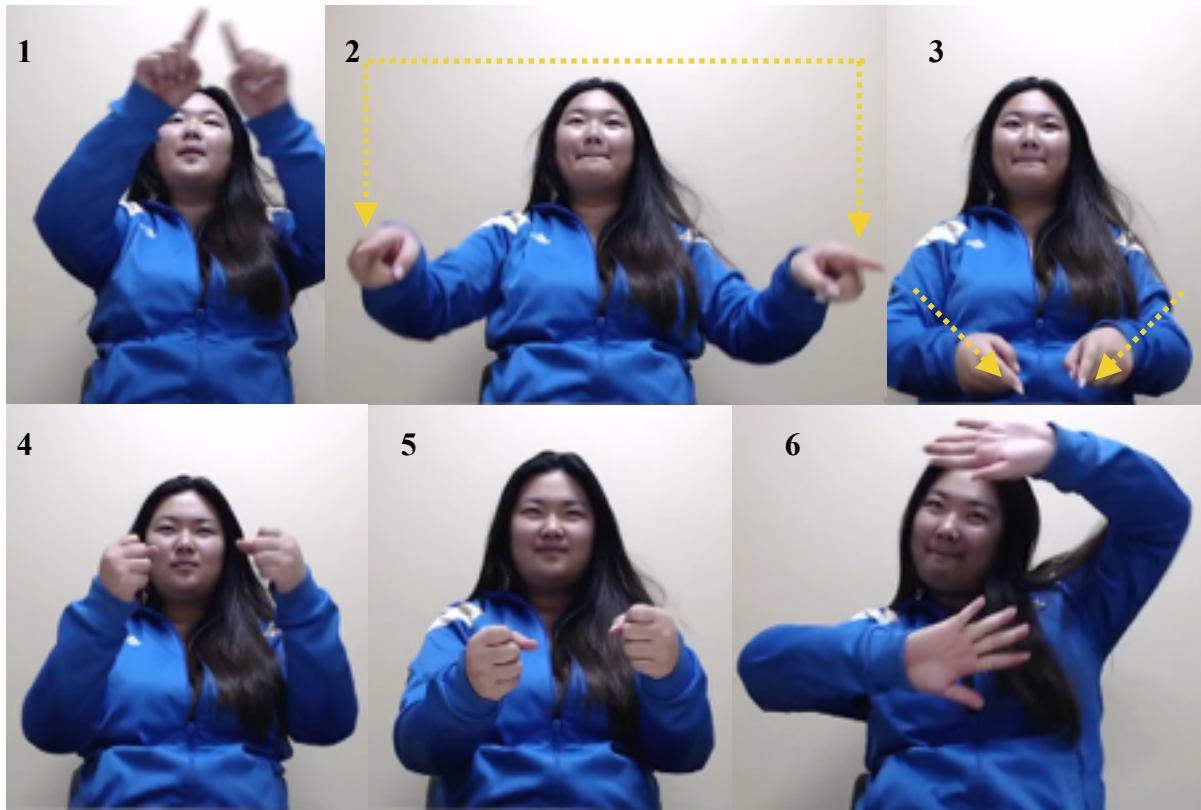
Image 4.8: Reformulation Repair: "A Shield" is changed from a traced-shape and handled gesture to an action gesture following misinterpretation.

Repair is initiated by either the Director themselves or by the Matcher. Other-initiated repair by the Matcher was instantiated in three manners: an incorrect guess (only in the Do-Over condition), the Matcher's gaze or facial expression, or the Matcher's prolonged guess time. Self-initiated repair, that is, by the Director, could be performed if the Director did not recall the grammatical form on the target card or if they believed their first gesture was not effective though no indication was given by the Matcher. Other-initiated repair was more common than self-initiated repair, especially in the Do-Over condition.

We hypothesized repair would manifest in supplemental gestures that are checks for understanding, or clarifications between the noun and verb forms. Nouns were repaired twice as often as verbs in the Do-Over condition, but were repaired in similar numbers in the Standard

condition. Furthermore, nouns were repaired most often through clarifications (three-forths of repaired nouns in the Do-Over condition and two-thirds of repaired nouns in the Standard condition), a repair strategy that often employs emphatic features in natural conversation, a characteristic that seems to arise in this experimental setting as well.

Looking to the most frequently used repair strategy, clarifications, a generalized logistic regression model was run to determine if the dependent variable, clarification repair, was predicted by generation. It was hypothesized that clarification repairs would decrease in use over generations, as the need for them as a negotiation tool would diminish as well. We found that generation did not predict repair, that is, repair does not increase or decrease significantly across generations (see Table 4.5).

	Standard Condition		Do-Over Condition	
	z value	*Pr(>|z|)*	*z value*	*Pr(>|z|)*
Generation 1	-9.678	<2e-16 ***	-9.388	<2e-16 ***
Generation 2	0.115	0.9087	-1.820	0.0687 .
Generation 3	1.293	0.1962	-1.849	0.0645 .
Generation 4	0.858	0.3908	-0.940	0.3475

Table 4.5: Clarification Repairs by Generation

Given the tendency to clarify through the addition of a marker, we used a generalized logistical regression model to determine if clarification use predicted marker use, as compared with no repair use at all. There was not a significant relationship between these two variables in the Standard condition, but the Do-Over condition saw a significant (p = .013*) association between clarification use and marker use (see Table 6).

STANDARD	Estimate	SE	z score	Pr(>|z|)
No Repair	0.537351	0.1031	0.211	.88e-07 ***
Clarification	0.1558	0.3683	0.423	0.672
DO-OVER	Estimate	SE	z score	Pr(>|z|)
No Repair	0.4692	0.1034	4.537	5.71e-06 ***
Clarification	1.0067	0.4053	2.484	0.013 *

Table 4.6: Clarification Use Predicting Marker Presence

Though, it should be noted that, marker use was still common without repair use. Considering the data together, repair use is not patterned, while most chains had a more patterned use of markers. The relationship between repair and marker use, then, may not be as strong as initially hypothesized.

3.4 Accuracy and Timing[1]

Overall, accuracy[2] increased from Generation 1 to Generation 4, though more substantially for the Standard Condition chains (see Figure 4.3). Accuracy peaked (i.e. high accuracy) most drastically from Generation 1 to Generation 2, in both conditions. We might attribute this trend to the role of the Observer, who comes into the interacting dyad as a knowing participant having seen the previous dyad perform the task.

---

[1] Gesture time and accuracy measures across chains include one chain in each condition that had a 5th generation; however, Generation 5 was not used in statistical models, as not all chains had this extension

[2] Accuracy is not considered a valid outcome measure for this experiment, as the chances to guess correctly technically increase over rounds within a generation, given the nature of the task (options are taken aways throughout the round). Nonetheless, some accuracy measures are considered here, but not in statistical models.
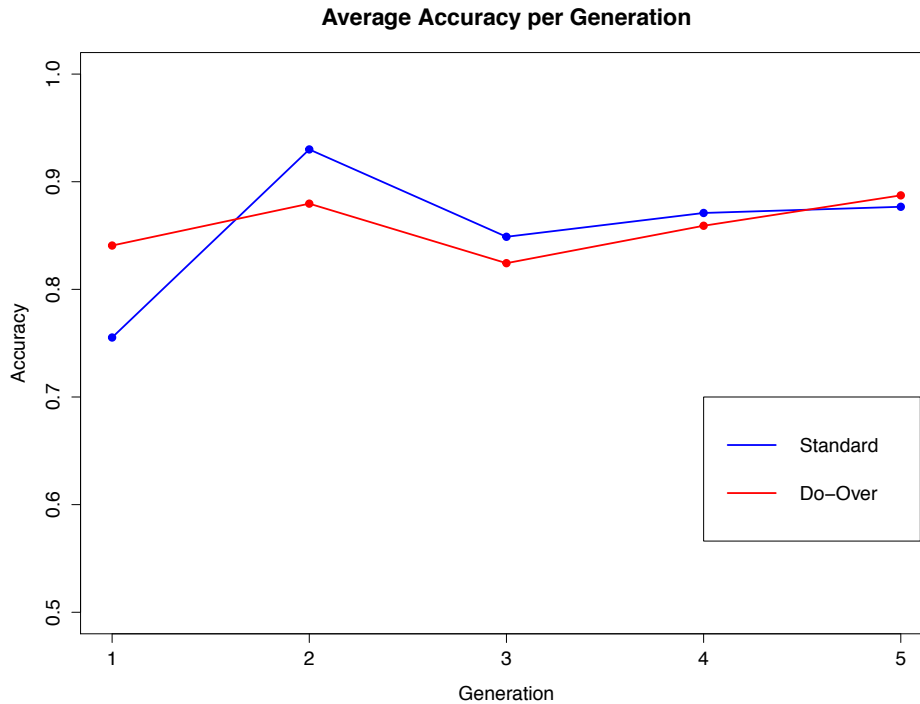
Figure 4.3: Average Accuracy by Generation

The increase in accuracy from generations 1 to 4 is significant for the Standard condition, but not for the Do-Over condition, as shown through a linear regression model determining correct guesses as predicted by generation (see Table 4.7). The Do-Over condition does, though, exhibit less change in accuracy over generations, which may be attributed to the ability to immediately repair a gesture given incorrect card choices. The ability to have an immediate do-over, often performed in seconds, could result in less pressure to be accurate in guessing since repair is not as time consuming. Furthermore, initial accuracy in the Do-Over condition was higher, leaving little room for improvement in subsequent generations.

	Standard Condition		Do-Over Condition	
	t value	*Pr(>|t|)*	*t value*	*Pr(>|t|)*
Gen1	32.972	< 2e-16 ***	35.708	<2e-16 ***
Gen2	5.160	3.04e-07 ***	1.156	0.248
Gen3	2.649	0.008220 **	-0.490	0.624
Gen4	3.492	0.000503 ***	0.548	0.584

Table 4.7: Accuracy by Generation, per Condition

Using the average accuracy of guesses, each condition demonstrated higher average accuracy when guessing noun-targets with the Index Finger marker (see Figure 4.4). The average accuracy using a the Index Finger marker was also higher than not using a marker at all; though, O-Emph and O-Point markers resulted in similar accuracy as not marking at all (with O-Point being slightly more advantageous).
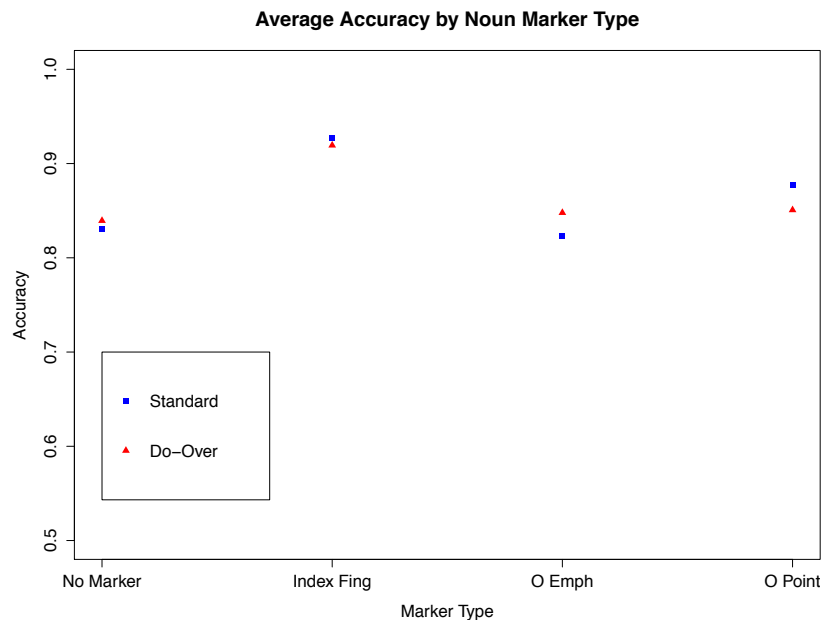
Figure 4.4: Accuracy in Marker Use: Collapsed by condition, the average accuracy for a given marker strategy, and no marker use.

For both conditions, a generalized logistic regression model was run to determine if the use of a marker predicted the accuracy of the guess (see Table 4.8). Only one noun-marker, the "Index

126

Finger," significantly predicted accuracy for both conditions (Standard: p = .0453*; and, Do-Over: p = .0246*). All other markers were not significantly associated with accuracy.

	Standard Condition		Do-Over Condition	
	z value	*Pr(>|z|)*	*z value*	*Pr(>|z|)*
No Marker	3.683	<2e-16 ***	15.300	<2e-16 ***
Index Fing	2.002	0.0453 *	2.248	0.0246 *
O Emph	-0.150	0.8807	0.207	0.8360
O Point	1.279	0.2011	0.242	0.8089

Table 4.8: Accuracy Predicted by Marker Use

The length of gestures did not vary greatly, with the exception of repair sequences. Overall, gestures were between 2 and 4 seconds long, though, on average Do-Over condition gestures were shorter than Standard condition ones for all generations (Figure 4.5).
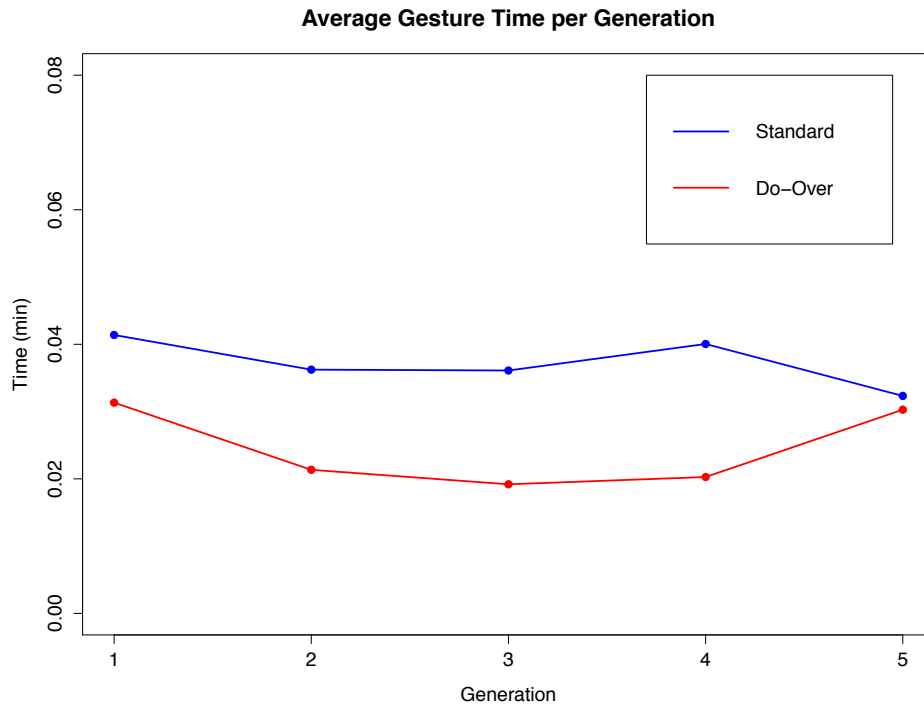
Figure 4.5: Average Gesture Length by Generation

127

While the difference in gesture length may seem minimal, a linear regression model (the independent variable was generation, and gesture length, in terms of time, was the dependent variable; see Table 4.9) revealed the decrease in time to be significant, expect for Generation 4 of the Standard condition, which saw an increase in gesture length, as did Generation 5 from the Do-Over condition.

	Standard Condition		DoOver Condition	
	t value	*Pr(>|t|)*	*t value*	*Pr(>|t|)*
Generation 1	1.050	< 2e-16 ***	3.609	< 2e-16 ***
Generation 2	-2.610	0.00920 **	-7.488	1.59e-13 ***
Generation 3	-2.567	0.01043 *	-9.167	< 2e-16 ***
Generation 4	-0.695	0.48753	-8.329	2.82e-16 ***
Generation 5*	-3.292	0.00103 **	-0.549	0.583

Table 4.9: Gesture Length Decrease Over Generations

Though the general trend of decreased gesture length, by seconds, appeared trivial, within the context of this study, minute changes have large effects. In general, as gestures were transmitted through generations, their length shortened, indicating the potential constricting of the gesture space and the diminished need to elaborate upon initial gestures.

DISCUSSION

The pilot study and previous experiments are indicative of the potential for repair to influence the structure of emerging communication systems, particularly in the gestural modality. Here we will analyze how repair and negotiation can facilitate conventionalization in the silent gesture system. Furthermore, looking to the system itself, we will examine how the system evolves over generations by building on prior strategies, and how this in turn leads to alignment and some evidence of compositionality.

Negotiation to Conventionalization: Repair for Noun Marking and Marker Order

As generations modify the previous generation's output, descriptive gestures become lexicalized such that an identifiable and systematic marking system develops in all chains. Typically, a marking gesture emerges from the need to disambiguate noun-verb pairs following evidence of inaccurate guess or confusion. Referring to the patterns illustrated in Figure 1 (in the Results), the Do-Over chains 1 and 2 see more rapid marker use, as opportunities for repair are instantaneous, wherein marking strategies emerge and become quite established by the end of Generation 1. Both these chains exhibit more clarification repairs in Generation 1 as well. Chain 3, however, used two marking strategies (Object Emphasis and Object Point) simultaneously in Generations 1 and 2, until Object Point became the preferred method in Generation 3. The increased use of clarification repairs coincided with the negotiation of one preferred marking systems. Fay et al (2010) similarly found repair (matcher feedback) to be present through Round 4 of their graphical communication game, but it decreased in use over rounds until it was nearly non-existent by Round 5. They posited that the need to repair had diminished over rounds as a global sign system took hold. In the present study, interestingly, marking systems may be arising due to the prevalence of clarification repairs, highlighting particular features of a given object or the object-ness of the target word.

Although the Standard condition chains do develop a marking system as well, they do so at a slower rate. Chains 1 and 2 used both Object Emphasis and Object Point marking, though by Generation 2 one form had become more prevalent. However, Chain 1 did witness a backsliding in the dominant system (Object Point). Chain 2 saw a dramatic shift in marker use from Generation 1 to 2, abandoning the O-Emph strategy for O-Point. Marker type switches may

result from a recognition of the guessing accuracy; for instance, in Chain 2, O-Emph markers only had a 71% accuracy rate, while Generation 2's introduction of O-Point co-occurred with a 92% accuracy rate (similar trends were observed in Standard Chain 3 and Do-over Chain 3). Both Chains 1 and 2 produced the least amount of clarifications overall, and Generation 4 was when the most clarifications were performed for these chains. Chain 3, on the other hand, experienced a definite switch in marking strategy halfway through the chain. The distinct switch from Object Emphasis and Object Point occurred in Generation 3 when one participant introduced the Index Finger marker. The new marker generation also performed the most clarifications within the chain. The Do-Over condition chains, in contrast had a significant association between clarification use and marker presence. An increased use in clarification repairs, then, may have an impact on the rapidity of conventionalized noun marker use. Overall, dominant marking strategies do not conventionalize until later generations in the Standard chains, which may be due to the inability to provide immediate repair to inefficient gestures.

The ubiquity of the marking systems in this experiment demonstrate its efficiency in communicating meanings, and, specifically, in disambiguating nouns from verbs. The immediate opportunity to modify the previous gesture (or gesture space) leads to repair-driven markers that conventionalize within a chain. Clarification repairs could be driving the gestural forms that are conventionalized; as the need to clarify a target meaning arises, and repair is performed, the strategy for repair becomes part of the gestural system. Here, the repair strategy "clarification" often results in an emphasis on the object-ness of the target word, including its handheld-ness, shape, or singularity. Marker clarifications are common, and their saliency could drive the marking system into conventionalization.

Marking became systematic not only in its use for nouns but also in its placement within the gesture itself. A fixed word order developed and became standardized to the point that participants would self-repair incorrect marker placement (order). Image 4.9 illustrates the self-repair of marking order, when the Director starts to gesture the action associated with "A Zipper" (still 1), but stops to hold up an Index Finger marker (still 2), and then continues to gesture the zipping action (still 3-4). The awareness of this word order, evidenced by the need to repair incorrect ordering, demonstrates its conventionalization. Late marking, resulting from incorrect placement, would often result in incorrect guesses by the Matcher.

Image 4.9: Marker Word Order Self-Repair: "A Zipper"

The Index Finger marker, as in the above example, was performed first in a gesture sequence, while the Object Emphasis marker was performed last in the sequence. That is, a Matcher would have to observe more of a gesture to know whether it was the noun or verb variant. Similarly, the Object Point marker was an insert gesture within the larger sequence. A Director would start with the action-variant of the noun and then would point to the location of the imagined object in the gesture space. After the point, the Director would often continue the action-variant aspect of the gesture (sometimes maintaining the point as well). While all three marker types were used to highlight noun-ness, only the Index Finger marker was a significant predictor of accuracy.

Given its placement in the gesture sequence, the Index Finger gesture appears to be more effective in noun-marking and accuracy in general. The immediacy of the marker in the sequence allows the Matcher to first narrow their selection to only nouns. Then, the task is only to match the action-variant gesture to the correct noun. This could remain a challenge, however, in instances of semantically similar nouns being in the Matcher's hand. For example, "a hammer" and "a nail" can be represented with the same action-variant (a hand acting as the hammer moving up and down, while the other hold a nail). In this case, the narrowing to a noun is no longer helpful in determining which noun is the target. For these situations we see three solutions: changes in orientation, handshape variations, or the addition of another marker after the original sequence. In the nail/hammer example, "a nail" might be held more precisely than in the "a hammer" gesture, while a change in orientation has the hammer gestured on a horizontal axis and the nail gestured on the vertical axis (see Image 4.10 below, for an example). Similar strategies were also employed to distinguish between semantically similar verbs. One participant commented on how they distinguished between similar words, stating "we did the same action

132

for hammering/nailing" (2C). Later generations would point to the noun-referent, co-opting a noun-marking strategy for the purpose of discriminating between the verbs.

Furthermore, participants remarked on the order in which they gestured nouns with the markers in the post-test questionnaire, demonstrating their awareness of the marking system and its order. In an Object Point dominated-chain, one participant commented that they "would wait to see if they [the Director] pointed to an object before repeating the action" (2D). This participant remarked on the need to "wait" to determine if the target is a noun or verb, potentially suggestive of its decreased efficiency compared to initial-position markers. In an Index Finger marking chain, another participant mentioned that she and her partner "would do the same gesture for nouns and verbs, but the noun, such as "a shovel," would start with us holding up an index finger to indicate 'one' or 'a'" (1Cr). The participant recognizes the first-position order for the Index Finger marker in noun gestures. A participant from an Object Point chain said "with nouns, we would end the action with gesturing to the noun's shape" (2Cr), though it should be noted it was not always shape as an emphasizer but also handheldness, as a participant from the same chain indicated, "if it was a noun, we would gesture that it was an object (usually by holding it up)" (2Br). Generally, it appears participants were consciously using the marking systems and understood how they were used in terms of application to nouns and wording in the gesture sequence. This awareness is indicative of a cultural convention of marking, how to use it, and what it communicates.

Repair as an Index for Conventionalization

The evolution of repair strategies over simulated time can be an index for the point of conventionalization the communication system has achieved. While the amount of repair does

not decrease with generation (as in Fay et al (2012), mentioned above), it may still be indicative of the conventionalization process in which repairs conform to prior preferences that are then re-constituted through the repair itself. There is evidence for the process of conventionalization in the repair sequences performed by the Directors themselves (self-repair) or those provided by the Matcher (other-repair).

Other-initiated repair is an attempt to get the Director to align with the Matcher's conceptions of the silent gesture system. When a Matcher initiates repair with facial gestures (e.g. furrowed brow) or a look back to the Director, they are communicating a misunderstanding of the prior, and attempting to make the Director repair to a more communicable gesture. If, on the other hand, the Matcher performs a repair (i.e. gives the repair options), they are demonstrating a more effective strategy which may then be taken up in subsequent gestures. Image 4.10 below is an example of other-initiated-other-repair (OIOR) in which the Matcher performs a repair on the gesture space through a point. The Matcher watches the Director perform the action associated with "A Nail" (Image 4.10, stills 1-2).

Image 4.10 (Stills 1-2): Other-Initiated-Other-Repair: Generation 1, "A Nail" initial gesture

After performing the action, the Director, uses a more precise handling handshape and traces the

nail shape in the air (stills 3a-4a). To this gesture, the Matcher points (still 4b) to the imagined

object that has been drawn in the gesture space as a means to clarify the intended meaning, "A

Nail," in contrast with "Nailing," another option in his hand. The Director nods in acceptance of

this repair. This chain continued to highlight the noun-ness through Object Emphasis to

distinguish between noun-verb pairs.

Image 4.10 (Stills 3-4): Other-Initiated-Other-Repair: Generation 1, "A Nail" Matcher repair

Another example of OIOR (Image 4.11) shows a Matcher giving options to the Director, which distinguish between the noun and verb forms of the prior gesture. The Director gestures "Whipping" with the same action three times (stills 1-2), but unbeknownst to her, the Matcher has both "Whipping" and "A Whip" as options for selection.
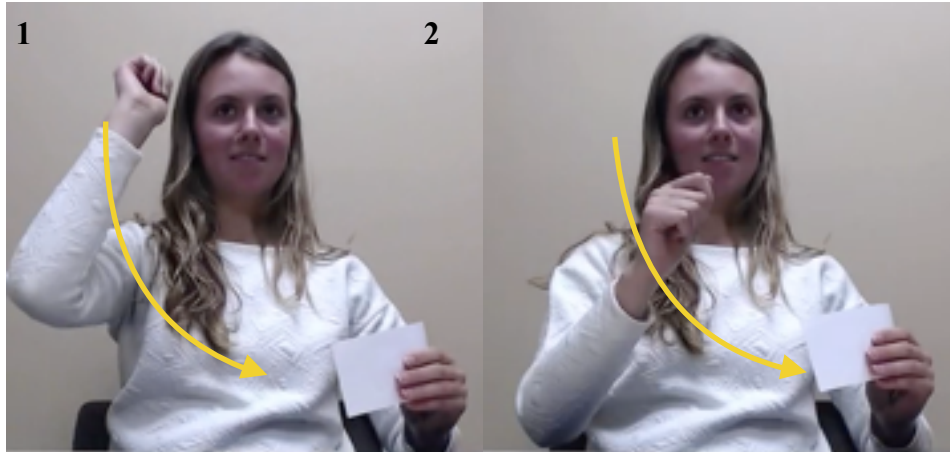
Image 4.11 (Stills 1-2): Other-Initiated-Other-Repair: Generation 1, Director, "Whipping" initial gesture

After the Director's gesture, the Matcher provides two options to clarify the prior gesture. First,

he re-performs the whipping action (still 3), and then points to the endpoint of the action (still 4),

to denote the action. Then, he holds the imagined whip in one hand, and points to the object with

the other (still 5), to denote the noun. This repair strategy allows the Director to be specific about

which meaning was intended by agreeing to one of the options provided.
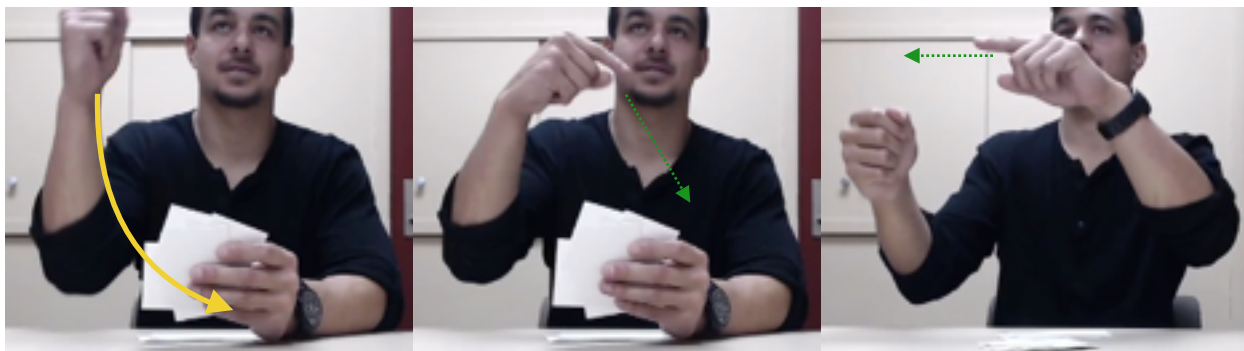
Image 4.11 (Stills 3-5): Other-Initiated-Other-Repair: Generation 1, Matcher repair, " Whipping"

Potentially resulting from the other-performed repair, this chain used more Object Point markers

in the following Generation, in which the same Director was still a participant. Repair, even from

the non-gesturing participant, can effectively facilitate the conventionalization of noun marking.

    Self-repair was particularly common when newly emerged marking systems were

misused in the task. Directors would overgeneralize a noun-marker to a verb, and would then be

required to repair the marker application- a testament to their knowledge of the emerging system. In Image 4.12, the Director has misappropriated the noun-marker Object Point to the verb "Phoning." While using an embodied "phone" gesture (still 1) she points to the object (still 2), but recovers the error by waving her hand to the Matcher (still 3) and re-doing the gesture to just the embodied phone (still 4). The pointing error may have arisen from the need to distinguish between "Phoning" and "Calling" (a distractor card), but having a system in place which dictates a point refers to a noun target, the misappropriation could have resulted in an incorrect selection.

Image 4.12: Self-Initiated-Self-Repair: Generation 2, " Phoning"

Due to their occurrence early in a chain, marker over generalizations can result in other innovative forms emerging. In the example below (Image 4.13), a Director has used the newly emerged Object Point when communicating "Ironing." To avoid having the Matcher select "An

138

Iron," as would be expected with the marker use, the Director self-repairs with an innovated verb-marking strategy. The Director attempts to communicate "Ironing" but points to the handled object (still 1). Realizing her mistake, she attempts to indicate the action is the referent (still 3), and also signals to the Matcher to disregard the prior gesture (stills 4). She then fans her hand over the entire gesture space (stills 5-6) to mark the whole space in which the action takes place is the intended referent, that it is a verb.
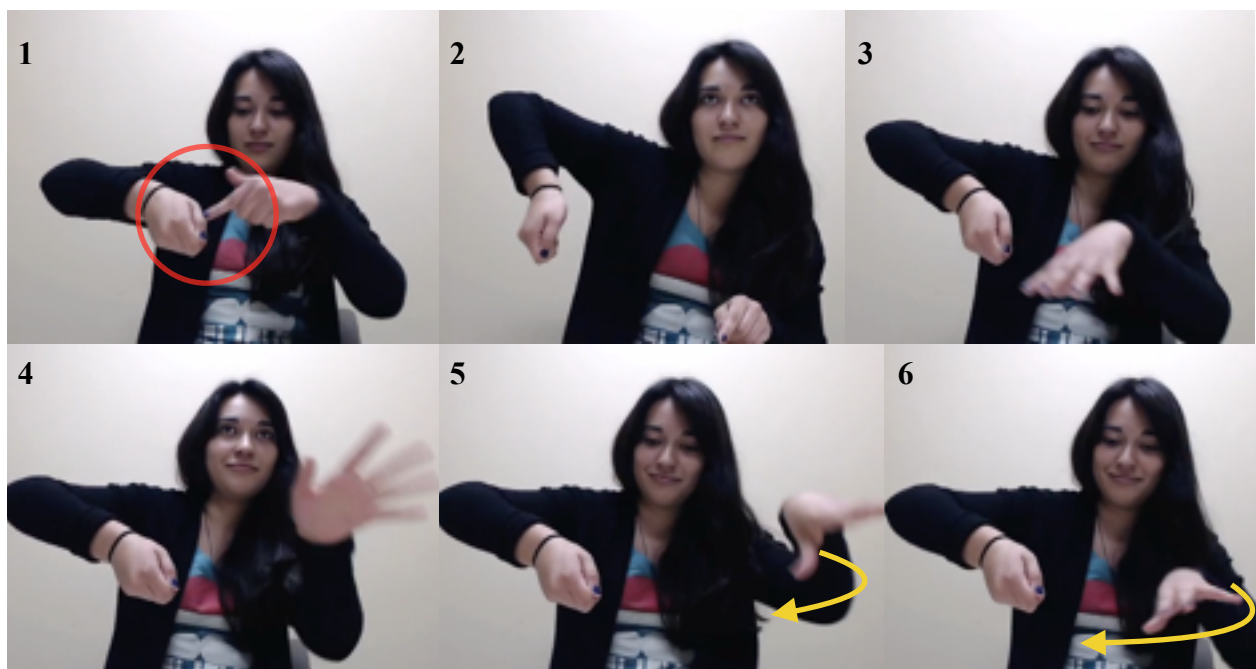
Image 4.13: Self-Initiated-Self-Repair: Generation 1, "Ironing"

The "Hand Fan" verb-marking strategy was present through the first two generations of this chain, but lost favor when a more accurate and salient noun-marking system, the Index Finger, emerged in Generation 3. These self-repairs on marker overgeneralization occur early in chains, and as they standardize, become less frequent. Again, self-repair in these situations evidences how participants recognize what strategy has conventionalized for which purpose.

Building from Priors: Alignment and Compositionality

The changes that occur to the silent gesture system in these experiments can be attributed to the interaction that takes place between the participants, the gestures they produce that are selected for modification, and the negotiation and repair that improves upon the system-in-progress.

Systematicity, achieved through the alignment of participants in terms of the gestures they produce (including handshape and markers), emerges differentially in the experiment. As previously discussed, the conventionalization of noun markers occurred earlier in the Do-Over condition than in the Standard condition. Participants commented on the alignment of their gestures with those of their partners in the post-test. One first generation participant noted her "partner rarely distinguished" (2Ar) between nouns and verbs, though she had used the Object Emphasis marker at times. In Do-Over Chain 2, there is a noticeable decrease in the use of noun marking, to which one participant said he "could not tell the difference between a noun and verb" (3Cr) since his partner did not distinguish between them. By Generation 4 of the same chain, Object Point became a prevalent marker. Here, systematicity developed slowly over generations.

In other chains, participants in the first generation were able to comment on the marking system they had helped establish; comments from later participants demonstrate the systematicity of the early marking system. In a Do-Over chain, the first participant mentioned they "used [their] index finger to indicate a noun versus a verb" and that index finger "refer[ed] to a single thing" (1Ar). Later participants answered similarly, noting they would "hold up an index finger" to represent a noun. In this particular chain, systematicity emerged quickly as a salient and effective marking strategy was passed down through the chain.

The face-to-face, contingent interaction of the experiment allows participants to build from one another's prior gestures as a means of developing systematicity over generations. Once a gesture has been performed, it is available for future use and manipulation. Immediate modification on a gesture can involve a reference to the gesture space or a particular element of the gesture. These references can be seen in the above examples of OIOR, as when a Matcher points to the gesture space of the Director. More removed reference to a prior gesture was also demonstrated by some participants (as Director) who would gesture to the Matcher's previous gesture space, as a means to indicate they would be communicating a similar item. However, even when a gesture is far removed- temporally- from the current gesture sequence, there is evidence that participants recall, reuse, and build upon priors to communicate similar or the same meaning (as in gestures from one generation to the next).

Over generations, participants modify and systematize prior gestures to conform to emergent conventions in the silent gesture system. This process involves reusing elements of prior gestures, even if temporally distant, to communicate a meaning, and it is complementary to the same phenomenon that happens in speech produced over multiple turns (Goodwin, 2013). In the example below (Image 4.14), the transformation of "A Guard" over four generations builds from one gesture to the next, incorporates similar-semantic gestures, and adds a marking system, resulting in a concise, communicative, and compositional gestural structure. Generation 1 (stills 1a-1c) performs "A Guard" by enacting the duty of a guard, standing in a rigid pose and looking from side to side. While the Index Finger marker has been used prior to this target word, it does not have patterned use early in the generation. Generation 2's "A Guard" builds from the previous generation, re-incorporating the rigid body posture, but adding a salute (stills 2a-2b).

141

The Index Finger marker has not been transferred to all nouns yet, as it is mostly applied to instruments initially.
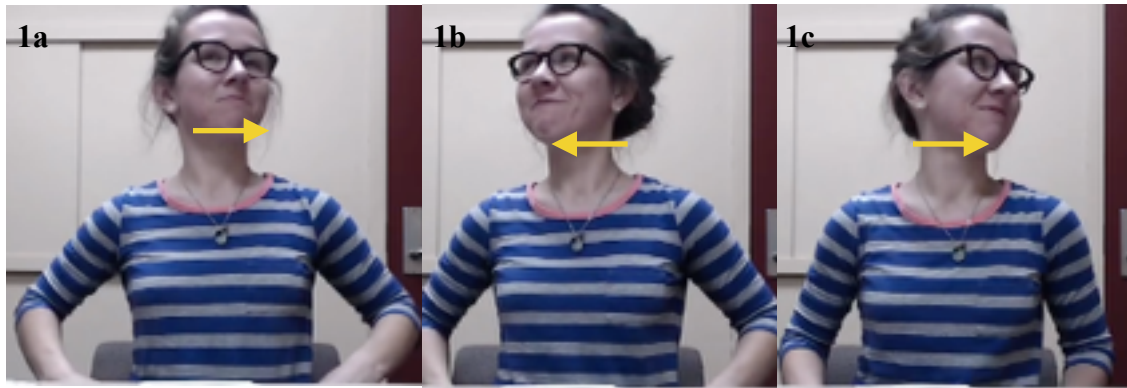
Image 4.14 (Stills 1a-1c): Transformations on Prior: Generation 1, "A Guard"

Image 4.14 (Stills 2a-2b): Transformations on Prior: Generation 2, "A Guard"

In Generation 3, the gesture for "A Shield," which does include the initial Index finger marker, produces a new structure wherein the shield is represented by crossed arms in front of the Director (stills 3a-3b). In the final generation, these elements are brought together in a systematic way to communicate "A Guard" by providing the Index Finger marker, then crossing the arms to indicate an action, and saluting to indicate the person-hood of the noun (stills 4a-4c).
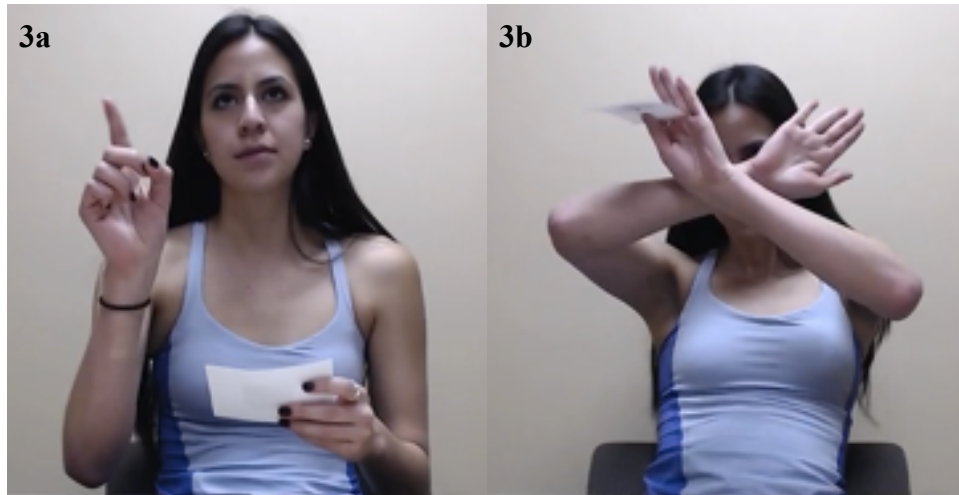
Image 4.14 (Stills 3a-3b): Transformations on Prior: Generation 3, "A Shield"
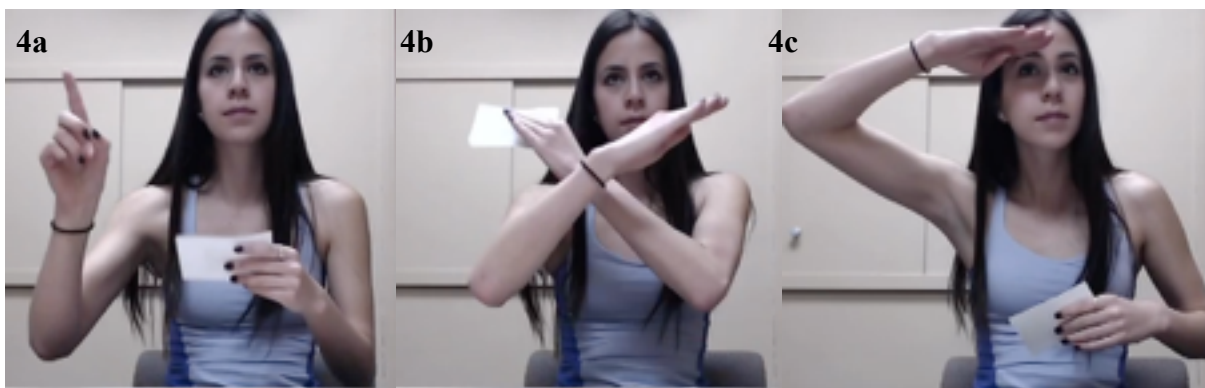
Image 4.14 (Stills 4a-4c): Transformations on Prior: Generation 4, "A Guard"

Building upon and transforming prior generations' gestures results in a systematic structure that is both expressive- it can distinguish between different meanings- and efficient- it can be easily transmitted and requires short, concise gestures to convey meaning.

Transformative operations on prior gestures can emerge through repair as well. The follow example (Image 4.15) emerged through the numerous repairs that were required of this pair to communicate "Picturing," when having to distinguish it from "A Picture" and "Photographing." Generation 2 used an "Action" (taking a picture) plus "Shape" (of the picture) handshape combination to communicate "Picturing" (stills 1a-1b) without difficulty. However,

143

Generation 3 had to distinguish "Picturing" from other semantically similar words, and therefore had to transform the prior gesture to one that more clearly discriminated it from other words. The Director uses a gesture from the eyes as if to look at something, and then draws her hands apart in front of her as if to display what is being "pictured" (stills 2a-2d). Generation 4 builds upon both gestures from prior generations to create a combined gesture that effectively communicates the target word, beginning with a "picturing taking" action and then repeating the same sequence from Generation 3 (stills 3a-3e).

Image 4.15 (Stills 1a-1b): Transformations on Prior: Generation 2, "Picturing"
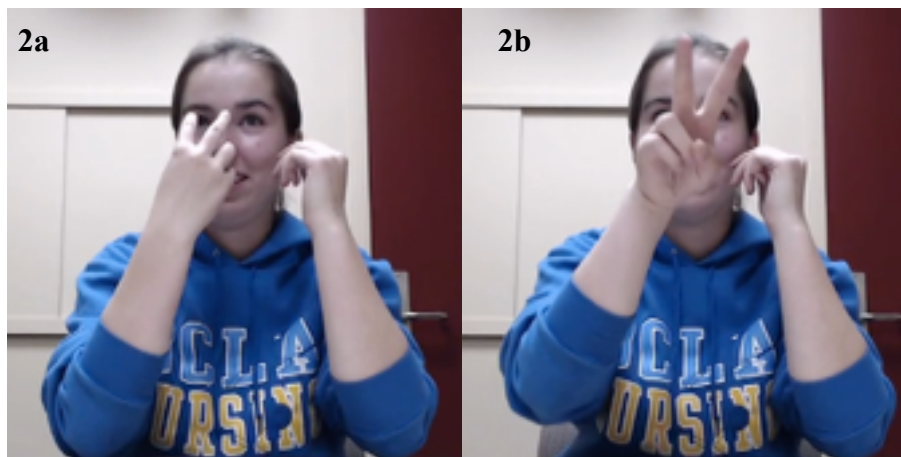
Image 4.15 (Stills 2a-2b): Transformations on Prior: Generation 3, "Picturing"

Image 4.15 (Stills 2c-2d): Transformations on Prior: Generation 3, "Picturing"
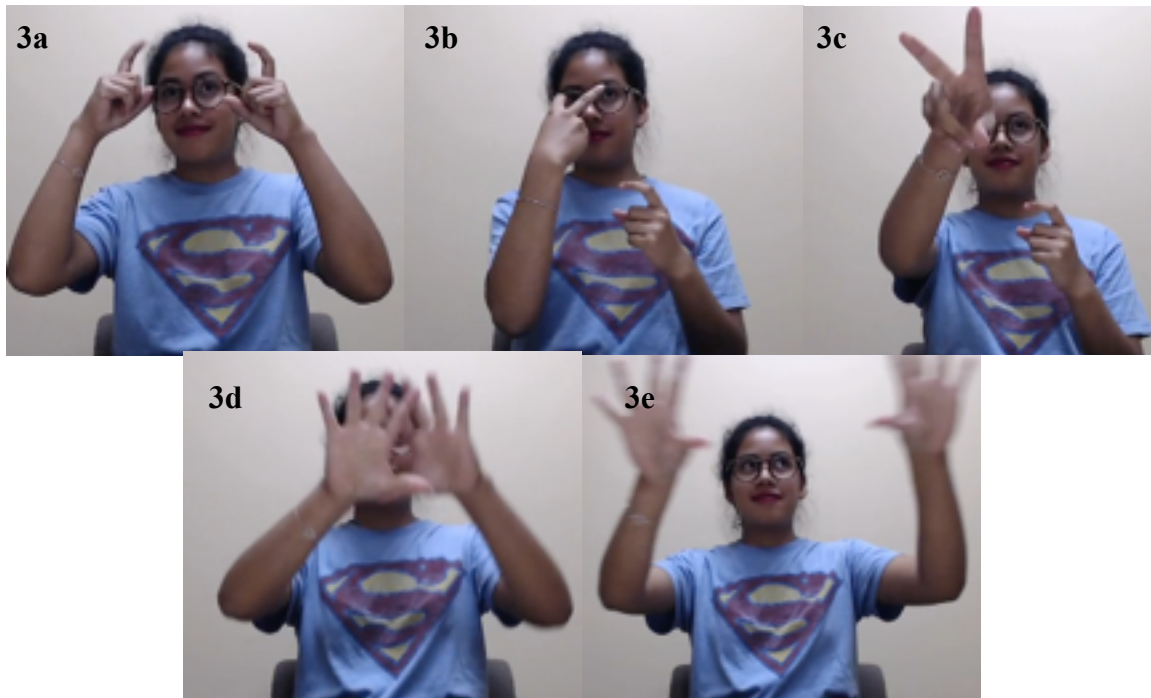
Image 4.15 (Stills 3a-3e): Transformations on Prior: Generation 4, "Picturing"

Just as interlocutors in conversation build from prior utterances that are no longer present, silent gesture users can build from prior gestures, making them more systematic over generations.

Compositionality, though not system-wide, evidences the ability of participants to build from priors to create different meanings. Compositionality, or composing meanings from individual semantic parts, is a seen in the noun marking system, but also in certain target words

that share morphological features, like "snow." Within a generation, participants built from the targets "Snow" and "Snowing" to compose meanings for "Snowboarding" and "A Snowboard."

"Snow" was gestured with articulated fingers moving downward (Image 4.16, still 1), while "Snowing" was similar yet the iconic gesture for "shivering" was incorporated (stills 2a-2b).

Image 4.16 (Stills 1-2b): Composition: "Snow" and "Snowing," respectively

Later in the generation, "Snowboarding" reused the "shivering" gestural element from "Snowing" with addition of the movement gesture for "downhill" (stills 3a-ab).

Image 4.16 (Stills 3a-3b): Composition: "Snowboarding"

Similarly, a Director communicating "A Snowboard" utilized the prior "Snowing" gesture, but first emphasized the shape of the object by spreading her hands across a flat area, leaving space

146

between her thumbs and fingers where an object might be (stills 4a-4b). Building from prior gestures can be a means from which to build compositional structure, as we see in this example.

Image 4.16 (Stills 4a-4d): Composition: "A Snowboard"

With a limited set of words, and few overlaps in meaning space, compositionality does not emerge system-wide, though its presence here may suggest our bias for it in communication systems. Follow-up surveys that extended the meaning space (to include potential compounds, and non-verbs and nouns) could give insights into if and how compositionality occurs as at the lexicon increases. Nevertheless, a morphological composition structure does emerge in this particular system, indicating the desire (and need) to make distinctions for expressive communication.

147

Languages of differing modalities across the world appear are characterized by design features (Hockett, 1960), some of which may distinguish human language from other non-human animal communication systems. Some of these features are duality of patterning (meaningless items combined to form meaningful items, and meaningful items combined to form meaningful items), arbitrariness, discreteness, semanticity, and traditional transmission (i.e. teaching and learning). As an emerging lab-language, the silent gesture system in this study only manifests some design features of language. For instance, teaching is not explicitly carried out in this experiment, but learning most certainly occurs, as with the adoption of noun markers, which are passed down to new generations. Furthermore, gestures themselves, as well as noun markers carry meaning, some iconic and some arbitrary (e.g. a raising index finger does not necessarily resemble a noun, particularly for non-count nouns as in "snow" and "milk" in the stimulus set). Given that participants in this study already have full language ability, looking at how they create a communication system that exhibits certain design features is the question we have sought to address here.

Similarly, newly emerging signed languages, as in Al-Sayyid Bedouin Sign Language (ABSL; Aronoff, Meir, Padden, & Sandler, 2008), currently employ only some design features of language, most notably not duality of patterning. Nonetheless, communication with ABSL is coherent among generations of users, who do have a conventionalized lexicon and build new compound forms from it. As with the current study, a conventionalized word order was also found in for modifiers. Furthermore, over generations (or, cohorts), gestured signs became smaller in size and shorter in length. This study, too, reveals a similar trend in the decrease of

148

gesture length (time of gestures). As a proxy for the natural observation of emerging signed language, experimental studies seem to replicate the evolution of systematicity and conventionalization that we observe in the real world.

The Case for Interaction

The crux of this study is to demonstrate the interaction, namely face-to-face contingent negotiation, has a role in the emergence and evolution of communication systems. Using an iterated learning paradigm, we can simulate language evolution over generations, but incorporating an additional feature- that of dyadic interaction with an observer- highlights the interactive component of language use and transmission. Both conditions followed this model, except that one sought to mimic natural conversation in terms of turn taking in repair sequences. Though some differences existed between the conditions, here we are concerned with the broader role of interaction in facilitating conventionalization.

The modified iterated learning model in this study derived from earlier studies of iterate learning of alien languages, though many did not provide contingent interaction scenarios. Kirby, Cornish, and Smith (2008) found that structure emerged between generations 4 and 9, with variability by chain. These interactions took place over a computer modulated communication, not face-to-face interaction. Nonetheless, a compositional structure emerged that allowed for motion, color, and shape to be encoded by different morphemes that could be strung together to encode meanings such as "black square moving in a circle." As noted, some chains attained their maximum level of structure by generation, while other did not until generation 9; often early-structure chains collapsed again into less-structured use of the language.

Face-to-face interaction, on the other hand, may promote quicker and more efficient transmission of the newly emerged language structure. Fay et al (2010)'s graphical communication task saw isolated pairs reduce drawing complexity through round 3 (an approximate to "generation 3"), however, community pairs took longer to achieve less complexity in drawings (note: simple drawing require less graphical refinement but still conveys meaning). The second condition here, though, is more akin to speakers of different languages- or dialects- coming together to communicate. The current study, in contrast, mimics the language learning environment of novices, including children and foreign language learners.

In a more naturalistic setting, Master, Schumann, and Sokolik's (1989) experimental creation of Persian and German pidgins required participants to extend their pidgin lexicon to new meanings. These systems displayed compositionality, regularity, and stabilized compound noun forms by their forth use. The rapidity of this systematization may have resulted from the extensive negotiation that took place when innovating new lexical items from a given set. In fact, just as with Fay et al, here noun compounds simplified by 66% (for example, an initial noun compound may have been comprised of 6 individual tokens, strung together, but after negotiation, decreased to about 3 tokens). It should be noted that this happened at the community level, as pairs exchanged partners in a gradual-turnover fashion, similar to the generational changes incorporated in our study.

Iterated learning models, as indicated by the name, are meant to resemble language learning and transmission, it is important to consider how acquisition of signed systems occurs, to make comparisons with the current findings. Ortega, Sumer, and Ozyurek (2012) investigated two possible variants that can occur in certain nominal tokens in Turkish Sign Language (*Turk*

150

Isaret Dili, TID): the perceptual-variant (e.g. a toothbrush is signed with the finger embodying the shape of the toothbrush and performing the associated action) and action-variant (e.g. a toothbrush is signed by holding an imagined toothbrush and performing the associated action). They found children of TID were more likely to use the action-variant, as were adults communicating with children (though the adult preferred strategy was the perceptual-variant). Action-variants, they suggest, may facilitate learning/language acquisition as they map more readily to iconic and motoric representations. Similarly, we find that participants in this study relied on action-variants for nouns (even when they were required to disambiguate noun and verb forms), perhaps as a means to bootstrap verb meanings to the noun. Then, to distinguish between the pair, participants innovated noun markers which conventionalized in nearly all chains. Learning, being a central feature to languages, can impact the trajectory and structure of communication systems.

In sum, face-to-face interaction might allow for more efficient and reliable transmission over generations. Interacting dyads, and groups to some extent, reach stabilization earlier than some non-interactive conditions, and may lead to the emergence and fixation of many features of systematicity simultaneously (being promoted by negotiation of varied aspects of meaning through moment-to-moment shifts and alignments). Teaching and learning are naturally interactive activities, therefore studies of the cultural transmission of language should consider how features of interaction impact the trajectory, in terms of evolving systematicity, of those languages.

CHAPTER 5

CONCLUSIONS ON INTERACTION AND THE EVOLUTION OF HUMAN

COMMUNICATION SYSTEMS

It is nearly impossible to identify what specifically allowed humans to have languages that are complex, abstract, and communicative. However, we are able to glean how languages acquire these features in laboratory experiments, and we can detail their functionality in naturalistic observations. New languages are emerging even today, a majority of them being signed languages in isolated communities where deafness has become quite prevalent. Cohorts of deaf individuals innovate a signed system that allows them to communicate among one another, and that, in turn, is passed down to younger generations of users. The young learners modify the system they have inherited to make it more concise, in terms of sign size and composition. Over generations, new emerging signed languages have evolved through the interactive use and transmission of the new system. Laboratory studies give us a more controlled look into the phenomena involved in interactive language learning and transmission so that we might better understand the processes that produce a communicative system that exhibits design features of complex human language.

SUMMARY OF EXPERIMENTAL FINDINGS

Though participants in these experiments are fully competent in language use (all are native or native-like speakers of English) and therefore have knowledge of language conventions, we can glean how certain features of language emerge and evolve across simulated generations by requiring the use of a novel communication form. While participants might rely on prior knowledge of language use, they will nevertheless be required to negotiate a new system

152

to be successful at the experimental task. Here, the emergence and evolution of a communicative system through innovation, negotiation, and conventionalization takes place in the interaction between individuals attempting to make form-meaning matches to improve their communication.

Study 1: Gestural Communication Game

In the innovative face-to-face, iterated gestural communication task, participants attempted to match gestural meanings to videos. The target communication would ideally involve both the manner and path movements of a ball in the stimulus videos. Participants either gestured these two-part meanings simultaneously (both manner and path together) or sequentially (manner, then path, for example); though, 81% of the meanings were gestured simultaneously. Over simulated generations, the participants' accuracy in judging meanings from gestures increased, as the diversity in gestures decreased. That is, gestures became more similar - in terms of handshape, hand orientation, and hand movement for a given meaning over generations. In addition, gesture length (in seconds) decreased over generations. Overall, the gesture length and space decreased, making communication more concise and more closely resembling features of signed language.

Given the face-to-face, contingent nature of the task, eye gaze and repair were two interactive features that were also considered for their impact on the gestural communication system. Each chain of participants (comprising multiple "generations") developed its own pattern of eye gaze for getting attention, monitoring, and confirming. However, when deviations from the set-in pattern occurred, it marked another action: repair. Repair was often other-initiated, that is, by the Matcher, when a misunderstanding arose in the interaction. An inserted look to the Director, often combined with raised brows or a confused face, would indicate the need to repair.

153

Matchers did not, however, perform repairs, leaving this work to the Director. The majority of repairs were clarification repairs, which highlighted or emphasized a feature of the meaning. As a result, certain gestural features became more salient through clarification repairs and were then maintained throughout the chain. Repair, then, played a role in the conventionalization of the gestural communication system.

Study 2: Card Selection Task

In a design similar to the above study, participants in the second study had to use gestural communication to play a card-based, closed-set "charades" game. The set included easily confusable noun-verb pairs (in a set of noun-verbs within a similar semantic category), which would have to be distinguished to succeed at the task. Over generations, participants developed gestures to disambiguate nouns and verbs, focusing especially on nouns. Marking systems were innovated and passed down generations, though not all chains used the same system. The three most common, and consistent, noun markers were "object point," "object emphasis," and use of the "index finger" held up like a "one." Some generations fluctuated between the use of markers, but eventually one strategy dominated. In addition, each marker conformed to its own word order which was also innovated and recognized by the participants. Nearly all chains made use of a noun marking system to disambiguate noun-verb pairs, and by the final generation, many chains used the noun marker reliably.

The participants' repair strategies also played a role in the conventionalization of noun marking systems. Clarification repairs were the most frequently used repair strategies, particularly with noun targets. In the Do-Over condition, the immediate "repair" turn allowed for more repair use, and in this condition we witness the use of a clarification eliciting the use of a

marker. Since clarifications require emphasizing or highlighting a previously gestured feature, greater saliency of those features emerges and becomes a substrate from which participants can build meanings, including compound meanings. Over generations, participants develop conventionalized marking systems from the emphasized forms. Repair, a feature of interaction, does facilitate the systematization of the gestures.

Through these experiments, we can witness the evolution of design features of human language, including systematicity, compositionality, arbitrariness, and expressivity. Although the participants in lab experiments have a fully developed language (and brain), and might be using resources from their experience with that language, having them interact in a new communicative form allows us to see how a communication system that conforms to human (perhaps more narrowly, English-speaking) biases might emerge and evolve through interaction.

INTERACTION'S INFLUENCE ON EVOLVING COMMUNICATION SYSTEMS

Cultural Transmission of Interaction Features

From this review of the lab studies, we see that features of the communication system are learned and then passed down generations; that is, gesture-meanings are culturally transmitted. Gestures corresponding to manner and path meanings, and verb and noun meanings (including a noun marker) are negotiated between participants, learned by current and new users, and transmitted through interaction and observation to new generations. However, the communication system itself is not the only feature of the interaction to be transmitted over generations. Eye gaze patterns of both the Director and Matcher role seemed to conventionalize over generations, and each lineage (in the first lab study) developed its own pattern of gaze. While some aspects of this pattern were influenced by the task (looking at the screen last, for

155

example), others may have been a result of the interaction and preferences. For example, a Director's gaze to the Matcher when beginning a gesture sequence can be interpreted as the desire to gain the Matcher's attention, as well as to monitor it; while the Director's initial gaze to their own gesture could be a means to establish joint attention to the gesture. Both these strategies can be helpful to the task and interaction, but preference may play a role in establishing which one dominates. Each strategy, though, did conventionalize as both Directors and Matchers of subsequent generations maintained the same pattern that had been established by prior generations.

Similarly, though every lineage (i.e. chain) in the second lab study developed a noun marking system, the strategies varied across groups. All chains negotiated marking systems over generations; some established one marker type early in the interaction, while others switched strategies as one became the preferred marker type. Though infrequent, some chains also attempted to mark verbs, but this marking system never exhibited the same success as the noun markers, and was not carried out beyond the early generations. Noun markers were varied, in placement ("word order") and in gestural form. From a hypothetically countless number of possible strategies, only three types were used by the participants. Furthermore, each lineage (of each condition) selected one of the three marker types as their dominant form (Index Fing, O-Emph, and O-Point). The variation in marker strategy reflects the diversity we find in natural languages around the world; many of the world's languages distinguish nouns from verbs, but do so in different manners. The same linguistic problem takes on numerous solutions, perhaps dependent on the needs, perceptions, and biases of the users themselves.

Just as languages and cultures worldwide exhibit variation in practices and conventions, so do the lineages in these experiments. As a micro-system, the experiment results reflect the variation we can find cross-linguistically. Although the strategies may differ for emergent morphological marking across chains, there is evidently a bias to innovate a systematic means to communicate certain types of meanings (or categories of meanings) and distinguish them from others. Though this is no doubt true of perhaps all languages, seeing how it emerges and evolves in a experimental language game can give us insight into what interaction can do to facilitate the conventionalization of design features of language, including systematic structure.

Finding Structure in Novel Systems: Negotiating Form-Meaning Fits

When playing a game of charades, there is no need to develop a systematic gesturing structure; instead, players use idiosyncratic gestures that conform to their own beliefs and thoughts about the target meaning. But, play the game over and over, with a closed set of meanings, with the same people, and you might see the development of a crude system of gestures. Perhaps this is based on the success of a particular gesturing strategy, or on individual's preferences, nonetheless, the beginnings of conventionalization occur. But, how does that arise? In face-to-face iterated learning games, we can observe how conventionalization and structure take hold in an emerging communication system.

Contingent interaction allows for the moment-to-moment monitoring and adjusting of one's own and others' gestural output. By monitoring the understanding of their partners, Directors can make adjustments to the gestures they perform; and, by indicating trouble (in the face, for example), Matchers give cues to their partners about the need to adjust the gestural output. Not having a rigid turn-taking structure- as in one facilitated by a computer (in which the

157

space bar is used to "give up" a turn to the interlocutor)- similarly allows for participants to interrupt the gesturer to request clarifications, in addition to providing an open interactive environment in which negotiation can take place fluidly. As such, negotiation and repair can happen immediately, and perhaps more effectively as some occur mid-gesture. Initiating a repair early can help disambiguate possible meanings, making guessing less effortful and more rapid. The repair sequence itself is also important to the negotiation of form-meaning fits.

To make the task more efficient, innovating a structure is key. However, partners must establish a structure and, without being able to discuss one, this occurs through emergent processes. Repair acts as a facilitator of this process in a number of ways. First, repair can signal when a gesture is confusing (or misunderstood), which could indicate it lacks comprehensibility. Repaired gestures, then, might include aspects that make them more understandable, and easier to match to target meanings. Second, repair can be performed to conform to previously preferred gestural strategies. Introducing a new gesture strategy (e.g. form, marker, order), can either provide more options for conventionalization or, more detrimentally, create misunderstanding. In the latter case, repairs back to the preferred format indicate the trajectory of the gestural system. The gesture (whether it be order, or handshape) becomes recognized for its preferred status, and is more likely to be maintained throughout the interaction; other gestures may become more like the preferred one, decreasing the diversity in the system and resulting in a conventionalized structure. Finally, repair- in particular, clarifications and repetitions- can increase the saliency of the gestured meaning. The more frequently, or more emphatically, a given gesture is performed, the more likely that gesture is to remain in the system. The negotiation within the interaction, largely carried out through repair, leads to alignment to a co-constructed gestural system that

158

exhibits structure in terms of handshape, movement, morphological marking, and compositionality.

Structure arises from the manipulation and re-use of prior gestural substrates, from which participants can refer to prior meanings and build compound meanings. The gestures produced by Directors are not permanent, yet participants can refer to them. To reference a prior gesture, participants pointed to the gesture space to disambiguate meanings (e.g. "a nail" versus " a hammer"). The gesture and the space in which it is performed provide a foundation on which participants can work to ensure understanding. However, this strategy is not just used with immediate priors, as some participants would refer to the previous turn (building on the previous Director's gesture in their own gesture), pointing or nodding to their co-participant or motioning to their gesture space. Drawing from this more distant substrate may require more cognitive effort, but might also highlight gestural features to become conventionalized; that is, the repetition of a gesture in two subsequent turns might increase its chances to become standardized. This can be seen in the re-incorporation of similar gestures, or parts of gestures, over generations and among novice users.

In addition, prior gestural substrates can provide the building blocks for compositional meanings. One compositional feature was the marker for "noun-ness" in the second lab experiment; anytime this morphological marker was included as part of the gesture, it was known to represent the noun-form of the target meaning. Participant posttest responses indicated their knowledge of the feature, including the order in which it appeared in the larger gestural structure. Most lineages saw marker use increase over generations, while some experienced shifts in marker use mid-chain. The introduction of the marker, usually as a means to repair prior

159

misinterpretation, allowed for structure to emerge in the system, in both composition and order. Compositionality also emerged in some compound target meanings, in which participants would build from a prior meaning ("snow," for example) for similar meanings ("a snowboard"). Substrate gestures, then, provided a baseline onto which additional meaning can be appended; in the given example, the gesture for "snow"- fingers moving downward mimicking snowfall- can be modified with a noun-marker ("Index Finger," for instance) plus the "snow" gesture followed by a gesture for the shape of the board to produce "a snowboard." Tracking the gestures that co-participants use and building from them is an interactive means to shape the structure of a communicative system.

Conventionalization: How Repair Drives an Emerging Language System

In natural conversation, repair functions as an indicator that a trouble-source is present in the speech and that there is a need to address it through modification or correction. Repair may take on an additional function in emerging languages: that of indicating the conventionalized state of an individual gesture, in addition to the systems as a whole.

The trends in repair use in the first experiment suggest that certain repair strategies might be used as a means to communicate conventionalized state of a gesture. As a system emerges and undergoes negotiation for meaning-form matches, repair should be a notable feature. As a system becomes "fixed" and the need to repair for negotiation purposes subsides, repair should become less frequent over time, but used when conventions are not met or there is a miscommunication. When conventions are not met, the repair may come in a reformulation- but one back to the standardized form. However, when a miscommunication takes place (including the Matcher not seeing the gesture), repetition should be the preferred repair format. All Generation 5 repairs

were repetitions, which suggests a diminished need for further reformulations and the potential for a fixed system. That is, repair acts as an index for the conventionalized state of the gestural system. The function of repair in emerging systems may be different than in established languages, with its evolution from a negation tool to a regulation device.

However, this pattern of repair use was not found in the second experiment, though the usefulness of repair was evident. The two conditions in the second study exhibited different trends in repair use over generations; the standard condition saw results similar to experiment one, but the do-over condition- allowing for immediate repair- had a decrease in repetition repairs over generations and a steady use of clarifications throughout a chain. These findings could be due to the task and stimulus set, which had a broader meaning space than the first experiment. We might eventually see similar trends in both experiments if experiment two were carried out over more generations. Though the trend differed in the do-over condition, the use of clarifications was associated with use of noun markers; that is, repair still had an effect on the system.

A RETURN TO DYNAMICAL/COMPLEX ADAPTIVE SYSTEMS

What the lab studies presented here and natural conversation data have in common is that they are based in a dynamical system of interacting and complementing resources that allow for meaningful communication to take place. As in natural conversation, participants in a face-to-face iterated learning communication game make use of eye gaze, facial expressions, turn-taking, and repair- sometimes simultaneously- to manage the task. Here, we return to the nature of complex adaptive systems (CASs), and how the experimental interactions in this dissertation exemplify the tenants of CASs.

161

1. "CASs are systems of complex structures in which patterns emerge dynamically through local interactions among many agents in spite of the absence of pre-ordained design" (Lee et al, 2009, pp. 17-18).

First, let's consider that "agents" could take on multiple meanings: actual human agents, and, more broadly, aspects of interaction (e.g. discourse features including eye gaze, facial expressions, and language- here, gesture). In the lab studies presented in this dissertation, we have purposefully allowed for human participants- the language users- to interact with one another in an attempt to foster negotiation and simulate a more natural language use setting. Furthermore, we have tested a new paradigm to include the gradual turn-over of participants from an observer role to an interacting role, mimicking language learning, transmission, and use (that is, novices learning from experts, and novices subsequently becoming experts). The local (generation and lineage-based) interactions of these agents allows for patterns to emerge in the silent gesture communication system.

Turning to the dynamic use of resources available to participants to generate patterned meanings, we can consider how Matchers' strategies work in concert with one another to highlight the need for a Director to repair the gesture. First, Matchers may deviate from a standardized pattern of eye gaze, in terms of locus and/or timing (a returned gaze to the Director before providing a guess, or a prolonged gaze to the screen or card set, respectively). In addition to gaze, Matchers will also indicate misunderstanding with facial gestures, namely a furrowed brow. Some Matchers will further highlight the request for repair with a point to the gesture space or even

162

provide an option for repair, using their own gesture. As Matchers call upon multiple resources to indicate the need for repair, Directors monitoring the Matcher's response will have multiple cues as to the appropriate next action: repair. Repair is brought about through the dynamic use of eye gaze, facial gestures, and the emerging language.

More broadly, the gestural communication system emerges from and evolves patterns through the interactive use of prior substrates, negotiation, and repair. These discourse structures, themselves arising from the pulling together of semiotic resources, work in concert to bring about structural change in the emerging language- or, at minimum, to make users aware of the potential for emergent structure. Manipulating prior gestures, determining successful strategies for that manipulation (adding markers, for instance), and repairing when miscommunication of the manipulated forms occurs all play a role in the development of a systematic structure to the new language. Regarding origins stories, why not posit that early humans made use of similar strategies to build a language- its lexicon and grammar? Co-operative and cooperative interaction in using a novel communication system can lead to its conventionalizations and systematization.

What might be problematic here is that "pre-ordained"-ness may not be absent entirely due to the the participants' native language. That is, they may be able to derive structure based on their own knowledge of and experience with language(s). Nevertheless, engaging in a modality of communication that is distinct from their native language does permit some absence of pre-ordained design; though co-speech

163

gesture may aid in the gesture use, it does not contain the grammaticality or systematicity that signed languages encode (see Perniss et al (2015) for a further review of the differences between co-speech gesture and signed languages).

2. "Small inputs into a CAS can cause major changes" (p. 18).

    Small inputs into a CAS work in concert with each other in non-linear interactions to produce large effects on the system. Repair, for example, provides slight modifications through clarifications, but the effects of this small input are amplified as participants build from the clarification, reincorporating them into new gestural forms, and negotiating the most comprehensive system, resulting in conventionalized forms.

    For example, the use of repair to introduce a noun marker eventually causes major changes to the gestural system. First, a new morphological form is added to the system, as a means to discriminate between easily confusable noun-verb pairs. Second, the addition of the marker requires the negotiation of where the marker will appear in a gestural sequence: initial, middle, or final position. Marker systems create an internal "word order" to which participants align. Repair, as a small input into the communication system, results in conventionalized forms that evolve from idiosyncratic ones.

3. CASs "show a general tendency for 'coherence under change'" (p. 18).

    While undergoing transformations (facilitated by negotiation and repair), the silent gesture systems in these lab studies demonstrate increased coherence over simulated generations of use to become more understandable, easier to produce, and

164

culturally transmittable. First, to make the silent gesture communication easier to understand, participants developed marker strategies that allowed for more efficient discrimination between noun-verb pairs in the second study. In both studies, but more reliably in the first one, participants aligned their handshape (in a representation of the ball at its movement) which decreases the cognitive effort to remember and recall numerous handshapes. In general, both systems could be considered understandable given the high accuracy of matching gestures to target meanings.

Languages that are easy to produce are essential for rapid communication, fitting a language to the capabilities of the human tongue, hands, and body while also considering the efficiency of the system. In both lab studies, the silent gestures became shorter in length (measured in seconds) over generations, suggesting that the gestured meanings became more concise over time. In the first experiment, the gesture space shrunk over generations; where early gestures were large, taking up much of the space in front of the gesturer, later gestures were much smaller. In conjunction with the increase in accuracy, shorter gestures still maintained semantic information sufficient for accurate guessing.

Finally, a language being learnable entails, in part, its ease to understand and produce, though other aspects of a structured system can promote its learnability. The experimental paradigm (iterated learning) simulates a learning environment in which transmission is both vertical (from "older" generations to "younger" generations) and horizontal (peer-to-peer transmission and learning within a generation). Having both transmission types in place, structure might emerge more rapidly allowing for

increased learnability. By generation 4, in both experiments, the silent gesture system conventionalized certain features that highlight systematicity. In the first experiment, this was demonstrated by the simultaneous gesture pattern as well as the decreased variation in handshape. For the second experiment, noun markers were nearly ubiquitous in chains, and handshape for nouns and verbs were typically quite similar with only the marker to distinguish them. Moreover, the second experimental language also had evidence of compound structure for nouns. Systematic structure allows for language to become more learnable as it provides a set of rules, or conventions, that could potentially be generalized beyond the closed set of meanings provided in the experiments. The silent gesture communication systems developed and conventionalized by interacting participants demonstrate "coherence under change" as participants' negotiation leads to the pruning of unnecessary and idiosyncratic features, while permitting the addition of systematic features that make language easier to learn and transmit culturally.

Just as isolated languages emerge and evolve through the bringing together of individuals, resources, and interactions, language itself likely emerges from a multitude of features in the biology, brain, social organization, and even genes of early humans. Complex systems have complex beginnings; to have a mode of communication able to express emotion, thought, past and future events, abstract ideas, and hypothetical scenarios with relative ease demonstrates the complexity language use entails. Our use of language is dynamic, we make meaning through dynamic systems interacting; should we not also consider that language emerges from dynamic processes?

FUTURE RESEARCH

Experimental studies allow for the manipulation of certain aspects of interaction; here the goal was to provide a set up in which human users were required to communicate with a novel system in order to learn how they might modify the system to become more usable, understandable, and learnable. Language is not used in a vacuum, and it is important to consider in the process of evolving a language, how that is done interactionally. As such, this new line of research draws from the actions and resources used in natural conversation, incorporating them into laboratory experiments to best imitate the real-world need for and use of language: in a context, with other individuals. We propose future research continue with real-world bases for experimentation, particularly for exploring language emergence and evolution.

Experimental studies following a similar paradigm as those presented in earlier chapters could still yield information on the emergence and evolution of communication systems. Meaning spaces (i.e. stimuli) could be expanded to include a variety of grammatical categories (e.g. adjectives) and more complex forms (e.g. sentences, as in "who did what to whom"). In addition, using participants from various linguistic backgrounds would be beneficial to determine if similar patterns for systematization occur given the native language of the participants; that is, how much impact is the native language having on the newly emerging one. It is important, too, to maintain the interactivity of the task to best mirror real-life language use. Interactive tasks such as instruction giving communicative tasks (e.g. a building task or map task) could provide more insights into how features of interaction contribute to the emergence and conventionalization of language systems.

Throughout this dissertation, we have shown how the complexities of interaction should not be disregarded when it comes to the study of language evolution. Natural conversation, even with early language learners, demonstrates how the dynamic use of language, body, and other resources allow for meanings to be conveyed, even in the face of incomplete language acquisition. Looking to the various origin theories of language, we see that it more plausible that there was more than one, singular 'cause' for language to emerge in our early ancestors. A combination of social, cognitive, and biological factors could have been at play, especially in the light of research on the co-evolution of culture and genes in language evolution. Finally, laboratory studies on the emergence and evolution of silent gesture communication have given evidence for the need to consider interaction as a pivotal aspect of language origins. Interaction among people is central to conventionalization, use, and the learning and transmission of language.

<u>Transcript 1</u>

```
01  GMA:    Quien se va a subir al tren,
02  GSN:    Baey-bie:[:s
03  GMA:               [Puedes contar los babies,
04  GSN:    Un-° ((mouths numbers 2 & 3, indiscriminate pointing to count objects))
05  GMA:    Cuantos son,
06          ( . )
07  GMA:    Hmm,
08  GMA:    ((GSN points to first count object)) U:no: ((GSN moves to next object))
09          ( . ) ((GSN moves hand to GMA's mouth, in hushing action))
10          ((GSN continues counting with 4 points while mouthing of numbers))
11  GMA:    Cuantos son,
12  GMA:    Dile a Christopher cuantos son
13  GSN:    ((Starts again with deictic point but silent counting))
14  GMA:    Cuentale en ingles w::one:=
15  GSN:    =w^w:on::e ((pointing to first count object)) ( . )   two° ((moves point with
16          each count object)) ((inaudible mouthing of numbers)) five° six° seven°
17  GMA:    Okay
```

```
01  GMA:    Who is getting on the train,
02  GSN:    Baey-bie:[:s
03  GMA:               [Can you count the babies,
04  GSN:    On-° ((mouths numbers 2 & 3, indiscriminate pointing to count objects))
05  GMA:    How many are they,
06          ( . )
07  GMA:    Hmm,
08  GMA:    ((GSN points to first count object)) O:n::e ((GSN moves to next object))
09          ( . ) ((GSN moves hand to GMA's mouth, in hushing action))
10          ((GSN continues counting with 4 points while mouthing of numbers))
11  GMA:    How many are they,
12  GMA:    Tell Christopher how many they are
13  GSN:    ((Starts again with deictic point but silent counting))
14  GMA:    Tell him in English w::one:=
15  GSN:    =w^w:on::e ((pointing to first count object)) ( . ) two° ((moves point with
16          each count object)) ((inaudible mouthing of numbers)) five° six° seven°
17  GMA:    Okay
```

Transcript 2

01 GSN: Mira ((point to count object in book)) (.) uno^ dos tres cuatro cinco seis oto
02 GMA: (Otra cosa) ((holds GSN hand in point to guide counting)) y uno:::
03 GMA: d[os   tres cuatro]
04 GSN: [dos tres cuatro] cin-
05 GMA: Ahora en ingles,
06 GMA: Quieres contar en ingles [o no
07 GSN: [no:::

01 GSN: Look ((point to count object in book)) (.) one^ two three four five six eight
02 GMA: (Another thing) ((holds GSN hand in point to guide counting)) and one:::
03 GMA: t[wo three four
04 GSN: [two three four fi-
05 GMA: Now in English,
06 GMA: Do you want to count in English [or no
07 GSN: [No:::

APPENDIX B: DIRECTOR EYE GAZE DURATION

Chain 1 Eye Gaze

Figure A1: Average Duration of Director Eye Gaze by Generation
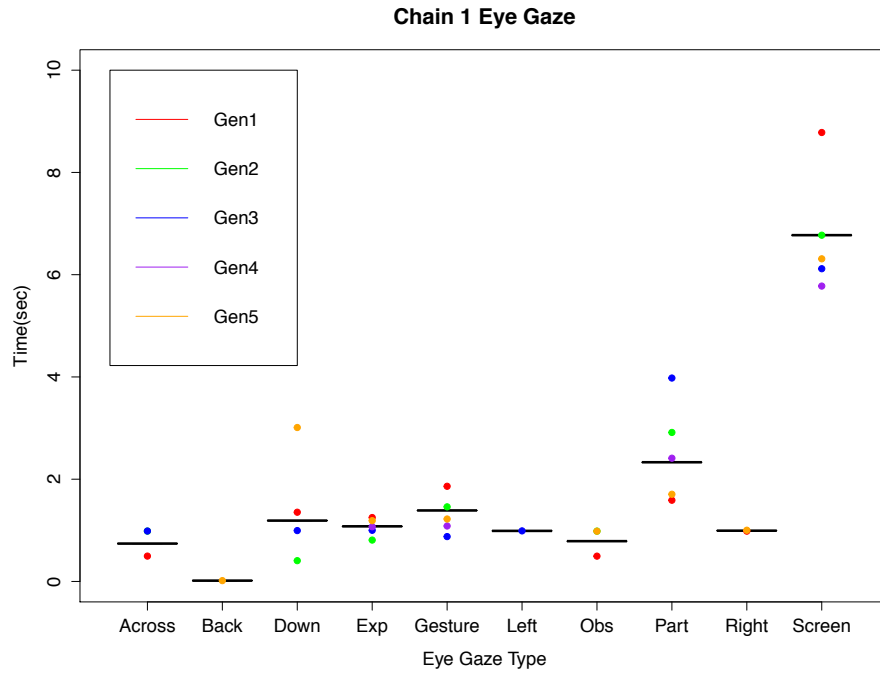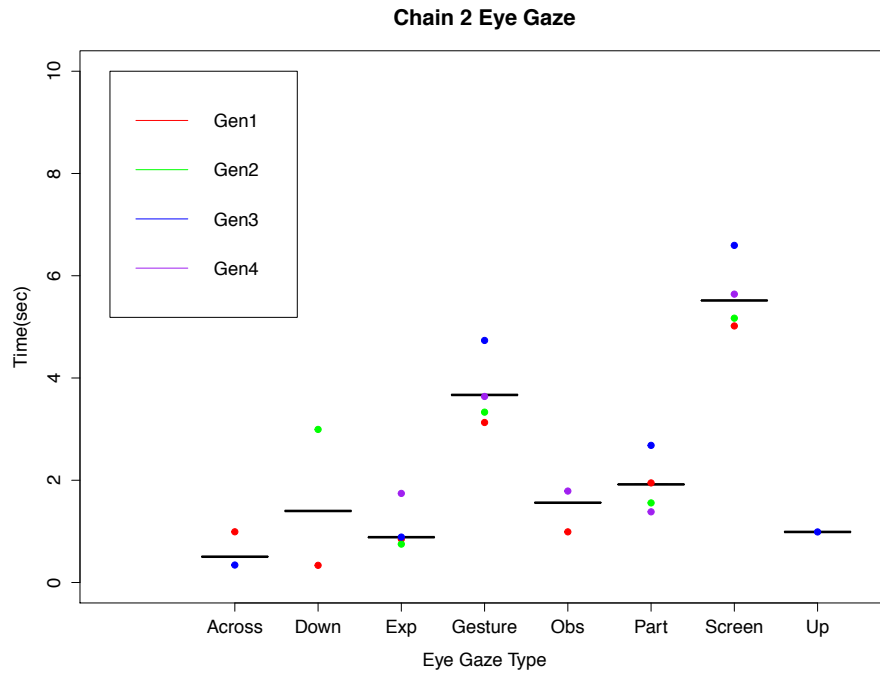
Chain 2 Eye Gaze

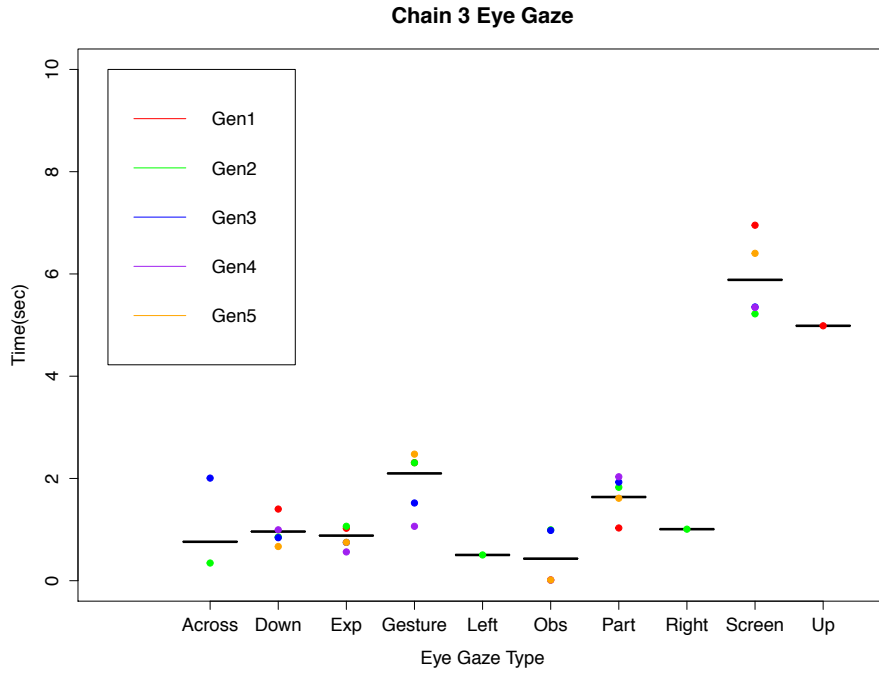**Figure A2: Average Duration of Director Eye Gaze by Generation**

171

Figure A3: Average Duration of Director Eye Gaze by Generation
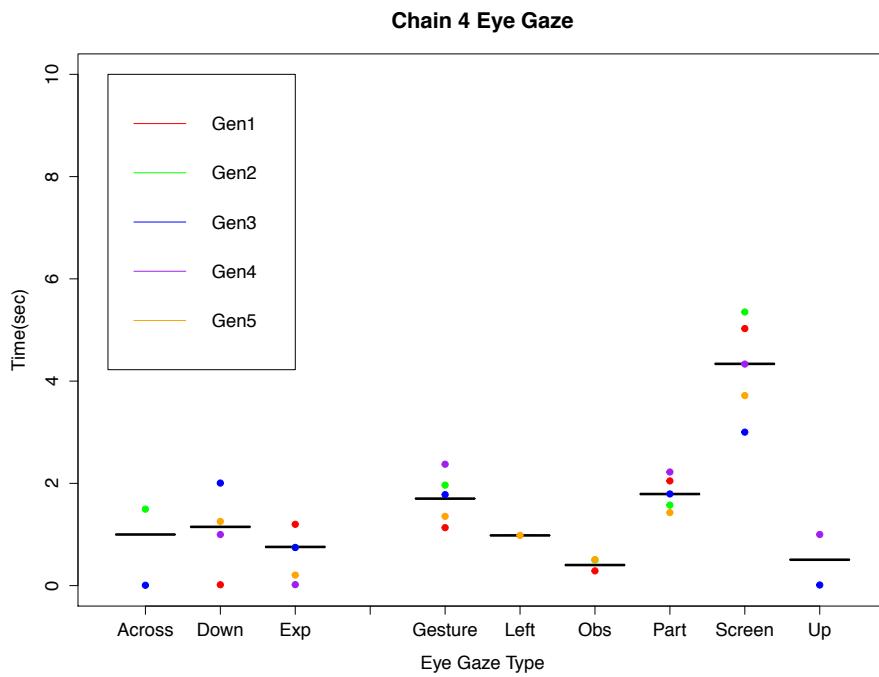

Figure A4: Average Duration of Director Eye Gaze by Generation

APPENDIX C: POST-TEST QUESTIONNAIRE

Interactive Card Selection Task
Post-Test Questionnaire

Please complete the following questions to the best of your ability, and relying on your
experience in the experimental setting.

A. Please rate on a scale of 1-5 (1= Strongly Disagree, 5= Strongly Agree) how much you agree
with the following comments based only on YOUR experience in the interactive card task.

1.	The task went smoothly (e.g. turn-taking, etc).	1	2	3	4	5
2.	You understood the goal of the task.	1	2	3	4	5
3.	You completed the task successfully.	1	2	3	4	5
4.	You believe your partner understood your meanings.	1	2	3	4	5
5.	You understood your partner's meanings.	1	2	3	4	5
6.	You knew when your partner wanted to make a correction.	1	2	3	4	5
7.	You could distinguish between similar cards.	1	2	3	4	5
8.	You could tell the similar cards were noun-verb pairs.	1	2	3	4	5

B. Please provide written responses with as much detail as necessary for the following
questions.

1. If applicable, describe how you knew your partner was making a correction or clarification.

2. If applicable, describe how you corrected or clarified your partner's meanings.

3. If applicable, describe how you and/or your partner distinguished between similar cards. If
you had different strategies, indicate that and describe both.

APPENDIX D: COMPARISON OF MARKER STRATEGIES WITHIN CHAINS

Generalized logistic regression models were run for each individual chain (given the variance in marker type use) in which generation (independent variable) predicted the difference in marker use frequency (dependent variable) of the two most frequently used marker types in a given chain (see Tables A1 and A2 for results on Standard and Do-Over condition chains, respectively).

Chain 1	O EMPH	O POINT					
			Factor	*Estimate*	*SE*	*z*	*Pr(>|z|)*
			Gen1	1.099E+00	6.667E-01	1.648	0.0994 .
			Gen2	3.168E-15	8.433E-01	0.000	1.0000
			Gen3	-9.163E-01	9.006E-01	-1.017	0.3090
			Gen4	-1.504E+00	9.28E-01	-1.621	0.1051
Chain 2	O EMPH	O POINT					
			Factor	*Estimate*	*SE*	*z*	*Pr(>|z|)*
			Gen1	-0.8650	0.4215	-2.052	0.040134 *
			Gen2	2.2513	0.6539	3.443	0.000576 ***
			Gen3	2.7368	0.8687	3.151	0.001629 **
			Gen4	3.1163	0.8545	3.647	0.000265 ***
Chain 3	INDEX FING	O POINT					
			Factor	*Estimate*	*SE*	*z*	*Pr(>|z|)*
			Gen1	2.057E+01	4.578E+03	0.004	0.996
			Gen2	8.402E-11	7.038E+03	0.000	1.000
			Gen3	-2.134E+01	4.578E+03	-0.005	0.996
			Gen4	-4.113e5.7	1E+04	-0.007	0.994

Table A1: Standard Condition Chains: Generational differences in frequency of marker user

Chain 1	Null	INDEX FING					
			Factor	Estimate	SE	z	Pr(>|z|)
			Gen1	0.5108	0.3266	1.564	0.11780
			Gen2	1.5041	0.6245	2.408	0.01602 *
			Gen3	2.2300	0.7993	2.790	0.00527 **
			Gen4	19.0552	1817.7600	0.010	0.99164
Chain 2	O EMPH	O POINT					
			Factor	Estimate	SE	z	Pr(>|z|)
			Gen1	-2.3514	0.7400	-3.177	0.00149 **
			Gen2	-0.6444	1.2639	-0.510	0.61018
			Gen3	-0.4818	1.2674	-0.380	0.70382
			Gen4	0.5596	0.9677	0.578	0.56308
Chain 3	O EMPH	O POINT					
			Factor	Estimate	SE	z	Pr(>|z|)
			Gen1	-0.1335	0.5175	-0.258	0.79640
			Gen2	16.6996	2399.5448	0.007	0.99445
			Gen3	0.8267	0.8763	0.943	0.34548
			Gen4	2.5314	0.9018	2.807	0.00500 **

Table A2: Do-Over Condition Chains: Generational differences in frequency of marker user

As a closer analysis, we calculated the predicted probability that specific marker strategies would be used over generations within individual chains (results in Table A3). The outputs below should be considered with Figure 1 (above) to contextualize the results. For example, that O-Point was predicted to be used 99% of the time in Generation 2 of Do-Over Chain 3, markers were used very infrequently in that particular generation- that is, nearly all markers were the O-Point variety, but not all nouns were marked. We do see, however, that the two chains (Standard Chain 3 and Do-Over Chain 1) which used Index Finger as the dominate marker were eventually predicted to use that marker regularly and exclusively by the final generation. Furthermore, Standard Chain 2 exhibits a dramatic shift in the predicted probability of the marker strategies, from O-Emph to O-Point in Generation 2.

STANDARD CHAIN 1	Index Fing	O-Emph	O-Point	DO-OVER CHAIN 1	Index Fing	O-Emph	O-Point
Generation 1	-	0.2500000	0.7500000	Generation 1	0.6250000	-	-
Generation 2	-	0.2500000	0.7500000	Generation 2	0.8823529	-	-
Generation 3	-	0.4545455	0.5454545	Generation 3	0.9393939	-	-
Generation 4	-	0.6000000	0.4000000	Generation 4	1.0000000	-	-
STANDARD CHAIN 2				DO-OVER CHAIN 2			
Generation 1	-	0.7037037	0.2962963	Generation 1	-	0.08695652	0.8521127
Generation 2	-	0.2000000	0.8000000	Generation 2	-	0.04761905	0.8521127
Generation 3	-	0.1333333	0.8666667	Generation 3	-	0.05555556	0.8521127
Generation 4	-	0.0952381	0.9047619	Generation 4	-	0.14285714	0.8521127
STANDARD CHAIN 3				DO-OVER CHAIN 3			
Generation 1	1.170226E-09	-	1.000000E+00	Generation 1	-	5.333333E-01	0.4666667
Generation 2	1.170226E-09	-	1.000000E+00	Generation 2	-	6.389220E-08	0.9999999
Generation 3	6.842105E-01	-	3.157895E-01	Generation 3	-	3.333333E-01	0.6666667
Generation 4	1.000000E+00	-	1.170226E-09	Generation 4	-	8.333333E-02	0.9166667

Table A3: Predicted Probabilities of Marker Strategy Use in Noun-Type Targets: Over Generations in Condition Chains (of two most used strategies per chain)

REFERENCES

Aronoff, M., Meir, I., Padden, C., & Sandler, W. (2008). The roots of linguistic organization in a new language. *Interaction studies, 9*(1), 133–153.

Bangerter, A., & Clark, H.H. (2003). Navigating joint projects with dialogue. *Cognitive Science , 27,* 195–225.

Beaken, M. (2011). *The making of language* (2nd ed.). Edinburgh, Scotland: Dunedin Academic Press Ltd.

Bickerton, D. (2000). How protolanguage became language. In C. Knight, M. Studdert-Kennedy, & J.R. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Functions and the Origins of Linguistic Form* (pp. 264-284). Cambridge: Cambridge University Press.

Bickerton, D. (2009). *Adam's tongue: How humans made language, how language made humans.* New York: Hill and Wang.

Botha, R.P. (2003). *Unravelling the evolution of language.* Language and Communication Library, 19. Oxford: Elsevier.

Boyd, R., & Richerson, P.J. (2005). *The origin and evolution of cultures*. New York: Oxford University Press.

Calvin, W.H., & Bickerton, D. (2000). *Lingua ex Machina:. Reconciling Darwin and Chomsky with the human brain*. Cambridge, MA: MIT Press.

Carpenter, M., & Call, J. (2013). How joint is the joint attention of apes and human infants? In H.S. Terrace & J. Metcalfe (Eds.), *Agency and Joint Attention*. New York: Oxford University Press.

Chater, N., & Christiansen, M.H. (2010). Language acquisition meets language evolution. *Cognitive Science*, *34*, 1131-1157.

Clark, E.V., & Clark, H.H. (1979). When nouns surface as verbs. *Language, 55*, 767–811.

Clark, H.H., & Krych, M.A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*(1), 62–81.

Corballis, M.C. (2010). Did language evolve before speech? In R.K. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Evolution: Biolinguistic Perspectives* (pp. 115-123). Cambridge: Cambridge University Press.

Cornish, H., Tamariz, M., & Kirby, S. (2009). Complex Adaptive Systems and the Origins of Adaptive Structure: What Experiments Can Tell Us. *Language Learning, (December)*, 187–205.

Csibra, G. & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of Change in Brain and Cognitive Development. Attention and Performance, XXI*, (pp. 249-274). Oxford: Oxford University Press.

Denton, A.W., Cosmides, L., Guemo, M., Robertson, T.E., & Tooby, J. (2012). The psycho-semantics of free-riding: Dissecting the architecture of a moral concept. HBES Conference. Albuquerque, NM.

Dessalles, J. (2000). Language and hominid politics. In C. Knight, M. Studdert-Kennedy, & J.R. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Functions and the Origins of Linguistic Form* (pp. 62-80). Cambridge: Cambridge University Press.

Donald, M. (1998). Mimesis and the Executive Suite: Missing links in language evolution. In J.R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the Evolution of Language* (pp. 44-67). Cambridge: Cambridge University Press.

Dor, D., & Jablonka, E. (2010). Plasticity and canalization in the evolution of linguistic communication: An evolutionary development approach. In R.K. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Evolution: Biolinguistic Perspectives* (pp. 135-147). Cambridge: Cambridge University Press.

Dunbar, R. (1998). Theory of mind and the evolution of language. In J.R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the Evolution of Language* (pp. 92-110). Cambridge: Cambridge University Press.

Everett, D. (2012). *Language: The cultural tool* (pp. 156-179). New York: Pantheon Books.

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

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

Galantucci, B. (2009). Experimental Semiotics: A New Approach for Studying Communication as a Form of Joint Action. *Topics in Cognitive Science, 1*(2), 393–410.

Gärdenfors, P., & Osvath, M. (2010). Prospection as a cognitive precursor to symbolic communication. In R.K. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Evolution: Biolinguistic Perspectives* (pp. 103-114). Cambridge: Cambridge University Press.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: Where might graphical systems come from? *Cognitive Science, 31* (6), 961-987.

Goldin-Meadow, S. et al (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America, 105* (27), 9163-9168.

Goodwin, C. (1980). Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning. *Sociological Inquiry, 50* (3-4), 272–302.

Goodwin, C. (2007). Environmentally coupled gestures. In S.D. Duncan, J. Cassell, & E.T. Levy (Eds.), *Gesture and the Dynamic Dimension of Language* (pp. 195-212). Amsterdam: John Benjamins.

Goodwin, C. (2013). The co-operative, transformative organization of human action and knowledge. *Journal of Pragmatics, 46*(1), 8-23.

Goodwin, C., & Goodwin, M.H. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica, 62* (1/2), 51-75.

Goodwin, C. & Goodwin, M. H. (1996) Seeing as a situated activity: Formulating planes. In Y. Engeström and D. Middleton (Eds.), *Cognition and Communication at Work* (pp. 61-95). Cambridge: Cambridge University Press.

Gould, S.J., & Vbra, E. (1982). Exaptation- A missing term in the science of form. *Paleobiology, 8*, 4-15.

Hawks, J. (2011). *The rise of humans: Great scientific debates*. Chantilly, VA: The Great Courses.

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

Hewlett, B.S. (2013). Early teaching and social learning in Aka foragers. *Brain, Culture, and Evolution Talk*. UCLA.

Hockett, C. F. (1960). The origin of speech. *Scientific American, 203*, 89–96.

Hurford, J. R. (2007). *The origins of meaning.* Oxford: Oxford University Press.

Hutchins, E. (1993). Learning to navigate. In S. Chaiklin and J. Lave (Eds.), *Understanding Practice: Perspectives on Activity and Context* (pp. 35-63). Cambridge: Cambridge University Press.

Jefferson, G. (1974). Error Correction as an Interactional Resource. *Language in Society 2*, 181-199.

Jefferson, G. (1983). On exposed and embedded correction in conversation. *Studium Linguistik, 14*, 58–68.

Kendon, A. (2009). Language's matrix. *Gesture, 9* (3), 355-372.

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

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America, 105* (31), 10681–6.

Kirby, S. & Smith, K. (unpublished). Cultural evolution of sign languages in the lab: From pantomime to linguistic systems. *Language Evolution and Cognition Talk Series.*

Lee, N., Mikesell, L., Joaquin, A.D.L., Mates, A.W., & Schumann, J.H. (2009). *The Interactional Instinct: The Evolution and Acquisition of Language.* Oxford: Oxford University Press.

Lieberman, P. (1991). *Uniquely human. The evolution of speech, thought, and selfless behavior.* Cambridge, MA: Harvard University Press.

Master, P., Schumann, J., & Sokolik, M.E. (1989). The experimental creation of a pidgin language. *Journal of Pidgin and Creole Languages, 4* (1), 37-63.

Ortega, G., Sümer, B., & Özyürek, A. (2012). Type of iconicity matters: Bias for action-based signs in sign language acquisition, 1114–1119.

Ozyürek, A., Furman, R., & Goldin-Meadow, S. (2015). On the way to language: Event segmentation in homesign and gesture. *Journal of child language, 42* (1), 64–94.

Perniss, P., Özyürek, A., & Morgan, G. (2015). The influence of the visual modality on language structure and conventionalization: Insights from sign language and gesture. *Topics in cognitive science, 7* (1), 2–11.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Brain and Behavioral Sciences, 13* (4), 707-784.

Power, C. (1998). Old wives' tales: The gossip hypothesis and the reliability of cheap signals. In J.R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the Evolution of Language* (pp. 111-129). Cambridge: Cambridge University Press.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696–735.

Schouwstra, M. (2012), *Semantic Structures, Communicative Strategies, and the Emergence of Language.* PhD Thesis, University of Amsterdam.

Schumann, J.H. (1997). *The Neurobiology of Affect in Language.* Malden, MA: Blackwell Publishers.

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. S. (2009). Signalling signalhood and the emergence of communication. *Cognition, 113*(2), 226–33.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Science, 7*(7), 308-312.

Steels, L. (2006). Experiments on the emergence of human communication. *Trends in cognitive sciences, 10* (8), 347–9.

Stivers, T., & Rossano, F. (2010). Mobilizing response. *Research on Language and Social Interaction, 43* (1), 3-31.

Stromswold, K. (2010). Genetics and the evolution of language: What genetic studies reveal about the evolution of language. In R.K. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Evolution: Biolinguistic Perspectives* (pp. 176-190). Cambridge: Cambridge University Press.

Tattersall, I. (2010). A putative role for language in the origin of human consciousness. In R.K. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Evolution: Biolinguistic Perspectives* (pp. 193-198). Cambridge: Cambridge University Press.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Phil. Trans. R. Soc. B, 364*, 2405-2415.

Tomasello, M. (2008). *Origins of human communication.* Cambridge, MA: The MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origin of cultural cognition. *Behavioral and Brain Sciences, 28*, 675-735.

Ulbaek, I. (1998). The origin of language and cognition. In J.R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the Evolution of Language* (pp. 30-43). Cambridge: Cambridge University Press.

Wilkins, W.k., & Wakefield, J. (1995). Brain evolution and neurolinguistic preconditions. *Brain and Behavioral Sciences, 18*, 161-182, 205-226.

Worden, R. (1998). The evolution of language from social intelligence. In J.R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the Evolution of Language* (pp. 148-168). Cambridge: Cambridge University Press.