THE UNIVERSITY *of* EDINBURGH

The evolution of symbolic communication: An embodied perspective

Jessica Erin Brown
BIS, MSc

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy
to
Linguistics and English Language
School of Philosophy, Psychology and Language Sciences
University of Edinburgh

12 September 2012

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Jessica Erin Brown

Acknowledgments

I thank my supervisors, Simon Kirby and Monica Tamariz, for their support and guidance throughout my time in Edinburgh. I am also extremely grateful to Eddie Dubourg for sharing his E-prime and other technological expertise. Special thanks is owed to Sean Roberts, whose patience, generosity and unparalleled acting skills made the experiments in this thesis possible.

I also wish to thank everyone in the Language Evolution and Computation Research Unit for comments, advice and camaraderie over the years. Thanks in particular to Chrissy Cuskley and Keelin Murray for keeping things light-hearted, and most especially to James Thomas for all the great company and conversation.

Finally, I am grateful to my parents for their limitless encouragement and support, and to my sister for keeping me sane and taking in a transient and her ornery cat.

Abstract

This thesis investigates the emergence in human evolution of communication through symbols, or conventional, arbitrary signs. Previous work has argued that symbolic speech was preceded by communication through nonarbitrary signs, but how vocal symbolic communication arose out of this has not been extensively studied. Thus far, past research has emphasized the advantages of vocal symbols and pointed to communicative and evolutionary pressures that would have spurred their development.

Based on semiotic principles, I examine emergence in terms of two factors underlying symbols: interpretation and conventionalization. I address the question with a consideration of embodied human experience – that is, accounting for the particular features that characterize human communication. This involves simultaneous expression through vocal and gestural modalities, each of which has distinct semiotic properties and serves distinct functions in language today. I examine research on emerging sign systems together with research on properties of human communication to address the question of symbol emergence in terms of the specific context of human evolution.

I argue that, instead of in response to pressures for improved communication, symbolic vocalizations could have emerged through blind cultural processes out of the conditions of multimodal nonarbitrary communication in place prior to modern language. Vocalizations would have been interpreted as arbitrary by virtue of their semiotic profile relative to that of gesture, and arbitrary vocalizations could have become conventionalized via the communicative support of nonarbitrary gestures. This scenario avoids appealing to improbable evolutionary and psychological processes and

provides a comprehensive and evolutionarily sound explanation for symbol emergence.

I present experiments that test hypotheses stemming from this claim. I show that novel arbitrary vocal forms are interpreted and adopted as symbols even when these are uninformative and gesture is the primary mode of communication. I also present computational models that simulate multi-channel, heterosemiotic communication like that of arbitrary speech and nonarbitrary gesture. These demonstrate that information like that provided by gesture can enable the conventionalization of symbols across a population. The results from experiments and simulations together support the claim that symbolic communication could arise naturally from multimodal nonarbitrary communication, offering an explanation for symbol emergence more consistent with evolutionary principles than existing proposals.

Contents

List of tables

List of figures

CHAPTER 1
Introduction

This thesis investigates the emergence of symbolic communication in human evolutionary history. A major step in the evolutionary process by which symbolic communication could have emerged has been proposed in the bodily mimesis hypothesis (Donald, 1991; Donald, 1998; Zlatev, Persson & Gärdenfors, 2005; Zlatev, 2007, 2008). The bodily mimesis hypothesis holds that prior to modern, symbolic language, humans evolved the ability to communicate through mimesis, or sensorimotor-based iconic and indexical signs. This ability provides a foundation from which symbolic communication could arise, but how such a transition would have taken place has not been fully examined. This thesis examines the gap between bodily mimesis and symbolic communication and aims to explain how that transition occurred. Thus, it will address the emergence of symbols from non-symbolic signs in human evolution.

The central question of this thesis is one about symbols, about human experience and about evolution. An investigation must therefore address all of these issues simultaneously and consider potential interactions. Each aspect of the question brings a set of ideas and constraints from its respective discipline that should be utilized and accounted for in the larger explanation. The present inquiry will interpret evidence and evaluate the validity of potential explanations guided by these factors.

The issue of human experience requires a consideration of the particular features of embodied human communication – that is, communication between beings with certain anatomical, cognitive and social attributes, as well as certain communicative demands. A recent multidisciplinary movement, broadly referred to as embodiment, recognizes the need to take into consideration the complexities and idiosyncrasies of lived experienced when attempting to understand a behavior or ability of an organism. An embodied perspective is necessary for both properly formulating a research question as well as identifying potential explanatory avenues for reaching an answer. Taking this perspective is crucial because it removes the question of this thesis from

the abstract and situates it in a specific context with a unique and causally relevant set of interacting forces.

The issue of evolution requires a consideration of evolutionary theory. Typically, evolution follows certain general patterns and principles, which can speak to the plausibility of a given scenario or outcome arising through evolutionary processes. These principles make it possible to assess the probability that certain events or series of events occurred in human evolution, and thereby formulate a viable evolutionary scenario for the emergence of symbols.

The issue of symbols requires a consideration of the nature of signs. Though signs as representational devices may at first seem straightforward and simple, capturing the relationship between form and meaning carried in a sign is more complex and conceptually nuanced than may appear on the surface. We can therefore carefully consider ideas from semiotics, which tell us how symbols are used to communicate and how to identify them. These principles are necessary for both evaluating research on symbols as well as formulating new hypotheses about their emergence.

This chapter will lay the explanatory framework that will guide this thesis. First, it will address the three aspects identified above: i) embodiment philosophy, ii) evolutionary principles and iii) the issue of symbolism itself. The ideas presented in this chapter will allow us to interpret evidence, evaluate the validity of potential explanations and formulate new solutions. Following that it will review the bodily mimesis hypothesis, which will serve as the theoretical backdrop from which this thesis will proceed. Finally, it will provide a brief overview of the research examined in this thesis and the arguments put forth.

1. Embodiment

Embodiment theory represents not one single discipline but an approach to explaining cognitive and physical phenomena that spans a number of disciplines. This section will give a brief overview of the central ideas behind the movement and then discuss how the perspective can inform this thesis.

1.1 Background and theory

Embodiment has its roots in the work of philosophers Edmond Husserl (1931, 1960), who examined the structure of experience and consciousness, and Maurice Merleau-Ponty (1962, 1963), who recognized the inextricable relationship between self and world and explored the nature and implications of this relationship. Building upon these ideas, Varela, Thompson & Rosch (1990) first proposed embodiment as a scientific framework. They argue that it is not possible to abstract away from embodied existence – encompassing bodily experiences, external cultural artifacts and social history – in trying to understand human abilities because it is precisely from these that behaviors and skills arise. That is, more generally, the specific physical, anatomical and social conditions of an organism and its environment are causally relevant to understanding the origin and underlying operations of its behaviors and abilities.

These ideas were later extended under what is called the enactive approach, which highlights the fact that a being's biological properties shape its experience in a very direct sense (Noë, 2004; Noë & O'Regan, 2001). This approach focuses on how the nature of our bodies determines how we act in and experience the world, and through these experiences our more sophisticated and complex behavioral and cognitive repertoires are built. The enactive position holds that it is both useful and necessary to consider the explanatory potential of lower-level, body-based factors in order to formulate an accurate understanding of how the abilities in these repertoires work. From this perspective, then, the most distinctive and remarkable abilities exhibited in human behavior may be explained in large part by looking to our embodied experience and the sorts of tools and scaffolding a rich physical and social world provides.

An embodied approach has proved theoretically fruitful and helped to solve notoriously difficult problems in cognitive science. The following section will describe examples that illustrate how embodiment, beyond providing a more realistic and comprehensive description of a phenomenon of interest, has explanatory power in itself.

1.2 Embodiment as a solution (not a problem)

Adequately considering the context in which human abilities arise does not merely result in a more valid description of the object under investigation. Doing so can serve simultaneously to discover solutions to seemingly intractable problems. As A. Clark (1997) notes, 'for many problems there is an elegant, often computationally and representationally low-cost solution that makes the most of gross physical properties of bodily platform and location situation' (pg. 205).

A. Clark (2008) cites as an example the fact that roboticists and engineers have spent substantial time and energy attempting to design and program robots to walk fluidly and flexibly. Though some success has been attained through the classical approach of engineering joints and programming movement sequences, such designs are far from energy efficient and do not cope well with novel environments. Designs that more closely resemble genuine human walking (in both appearance and energy requirements) are ones that make use of 'passive dynamics'. These designs exploit mechanical couplings between properties of the walking mechanism and the environment in which movement takes place. Power and motion are driven by the effects of gravity on the structure and weight of a design, mirroring human skeletal mechanics. This paradigm suggests that human-like locomotion can in large part be attributed to basic morphology (Collins, Wisse & Ruina, 2001; cited in A. Clark, 2008). In other words, the supposed computationally elaborate and complex act of walking may in fact simply 'fall out from' the physical relationship between anatomical features of the human body and its environment of movement.

Another example demonstrates how task-oriented skills can arise through the interaction between simple learning mechanisms and activity in the physical world. A. Clark (1997) describes a robot that was assigned the goal of collecting discarded soda cans in a busy laboratory – a task that required navigating a dynamic environment and identifying and collecting objects of interest. Though far beyond the scope of any contemporary classically programmed design, this goal was achieved by outfitting the robot with low-level 'sensory' systems to explore its environment and building in the potential for emergent coordination between these systems. The environment itself provided the structure and knowledge needed for success via the

robot's ability to sense and move within it. This research demonstrates that sophisticated behaviors can arise through simple mechanisms exploiting potential structuring that comes along with environmental input gained through active exploration.

Noë's (2004) in-depth exploration of the visual system also illustrates the principle that an embodied perspective is necessary and useful for understanding our abilities. The traditional view of visual perception conceived of the process as computations within the brain on stored internal representations of the environment, which in practice would entail vast amounts of storage for these representations and extensive, complex computations carried out on them. In contrast, Noë argues that if we consider the perceiver as a whole – an embodied animal that can interact with and gain information from its environment – a solution becomes available that makes more sense in terms of engineering and evolutionary constraints. That is, because the environment is immediately available to an active, embodied perceiver, the supposed role of internal representations can in large part be off-loaded on to that perceiver's ability to flexibly engage the environment to acquire the information needed at the time of seeing. Noë's approach provides a more comprehensive explanation of how vision operates, is computationally efficient, and is consistent with research on how visual perception is affected under various atypical experimental conditions. It is by integrating factors related to embodied experience – the kinds of bodies perceivers possess and the ways in which they act in the world – that this solution becomes available.

Finally, Thompson (2007) addresses the notoriously problematic concept of consciousness from the perspective of embodiment philosophy. He draws parallels between the way all living organisms derive meaning from the world through sensory explorations in the quest to survive and the similar relationship between mind and world. Thompson notes that self-sustaining beings, even at the level of single cells, are entities consisting of a dynamic relationship between their bodies and the world, and the nature of those bodies determines what in the world is meaningful to those beings. In the same vein, consciousness can be understood as dynamic interactions between beings with certain bodily and cognitive attributes and the environment – physical and social – in which they sense and act. By conceiving of consciousness

and mind as things that are carried out by whole, embodied beings in their environments – as opposed to internal to the brain – Thompson demystifies the concept of consciousness and lays a solid foundation for further understanding how it emerges.

1.3 Summary and discussion

The ideas and research reviewed here demonstrate the importance of accounting for lived experience in investigations of human traits. First and foremost, doing so ensures the object of investigation is comprehensively and accurately characterized, which is a prerequisite for understanding its origin and development. In addition, however, this approach often leads to the incorporation of factors that can in themselves inform an explanation. Thus, accounting for the particular features of situated organisms in their natural contexts is vital for an investigation of symbolic communication *as it is practiced by humans*.

Aspects of embodied experience are often thought of as 'messy' distracters that can be abstracted away. While this strategy may be useful and necessary depending on the level of analysis one takes as a starting point, the robotics research reviewed above shows we should not disregard the explanatory potential of context and embodied factors. Indeed, when the topic of inquiry is something as context-sensitive, interactive and dynamic as human communication, the embodied perspective becomes especially relevant.

The examples described above focus on how sensorimotor skills and interaction in the world construct higher cognitive skills in an individual. While symbolic communication is a trait that transcends the individual and exists on the cultural level, it is ultimately enacted, and presumably constructed, by the actions of individual, embodied humans. Thus, despite the second-order nature of our question the principles of embodiment are still applicable and relevant and should therefore be incorporated into the present investigation.

Some points that warrant brief mention concern ambiguities surrounding the notion of 'embodiment'. The word is often blanketly used to refer to what are in fact distinct

concepts, though these are of course related. For example, Wilson (2002) analyses theories concerning 'embodied cognition' and claims that six different categories of ideas are argued for under the same term. Embodiment can be used to talk about organisms on a biological, experiential or social level – or a combination of two or more of those levels. This subject of this thesis encompasses all three of these levels, as it focuses on the physiological features with which humans communicate, how communicative acts are perceived and how those are shared culturally. For present purposes, the notion of embodiment in these senses will be employed to accurately characterize communication as it is accomplished by humans, which will lay the explanatory groundwork for investigating the evolution of symbolic communication.

Another ambiguity in terminology to address is that surrounding the concept of 'multimodality'. Multimodality can refer to three different meanings: i) the combination of multiple sensory modalities, like auditory and visual communication; ii) the combination of multiple channels of communication, like gesture and text; and iii) the combination of different semiotic media, like spoken language and music. A given instance of multimodality can include more than one of these senses, but one does not necessarily entail another. For example, in theory communication could take place simultaneously through sign language and written language (imagine if a deaf signer were giving a presentation using slides with text), in which case the second sense holds but not the first. Similarly, song involves two different semiotic resources – language and music – but not two different modalities. To clarify the notion of multimodality in this thesis, we can note that it encompasses the first and second meanings, as it examines communication through two different channels that correspond to two different sensory modalities – vocal and gestural.

Based on the ideas from embodiment theory presented here, the question of symbol emergence can be framed in terms of the particular anatomical, perceptual and social factors characteristic of human experience. Specifically, it requires an examination of human communication in action and the modalities through which meaning is expressed. These factors will influence semiotic issues regarding sign use (discussed below in section 3), and taking an embodiment perspective therefore makes it possible

to address properly the question of symbol emergence in the context of human evolution.

2. Evolutionary theory

Research on evolution in various disciplines has established that certain principles typically govern evolutionary trajectories. These basic principles can therefore be used to evaluate the viability of hypothetical scenarios in human evolution proposed to explain the emergence of symbolic communication. This section will review these principles and discuss how they can inform the central question of this thesis.

2.1 Parsimony

The principle of parsimony is based on the idea that evolution typically works as a minimalist, with change being rare and taking place through the simplest and fewest available routes as possible (Sterelny & Griffiths, 1999). Thus, given equal evidence for two or more possible evolutionary scenarios, the one that involves the fewest transitions is the sequence with the greatest likelihood of reflecting the actual course of events. In other words, overly complicated scenarios involving events that are not strictly necessary to reach the known end state violate the principle of parsimony and should be assumed to have a low probability of having occurred in actual evolutionary history.

2.2 Continuity

Continuity is related to parsimony, in that new traits tend to be built upon existing ones as opposed to constructed from scratch. This is because co-opting and tweaking precedents requires fewer and more minor changes than creating entirely new structures to perform the same function. The concept of continuity also emphasizes the contingency-based nature of evolution (Gould, 1996). Evolutionary trajectories are contingent on the previous evolutionary history of an organism (Sterelny & Griffiths, 1999), meaning the current state of an organism limits and determines possible future avenues through which evolutionary change can occur. Continuity is based on the fact that an organism's particular physical and social characteristics

select which features of the environment are relevant to it and thereby guide and constrain evolutionary outcomes (Thompson, 2007). Thus, when evaluating or formulating scenarios, it is important to consider how traits that are known or presumed to be already in place could give rise to the trait under investigation.

2.3 Biological and cultural evolution

Thus far this discussion has spoken of evolution in a general sense; however, the issue in question here in fact involves two kinds of evolution: biological and cultural. The relative roles of biological and cultural evolution in leading to human symbolic communication as it exists today will be addressed in later chapters. For now it is only necessary to discuss their dynamics generally, which will provide the framework for assessing potential evolutionary explanations.

Biological and cultural evolution are separate – although highly interdependent (Richerson & Boyd, 2005) – processes that operate on distinct levels and targets. Broadly, the materials of biological evolution are genes and physical traits, while that of cultural evolution are cultural artifacts like tools, words and ideas. Biological evolution takes place typically on a timescale of many generations, while cultural evolution proceeds much more rapidly, with many changes often occurring in the lifetime of an individual. Despite these core differences, there is strong evidence that the two share deep similarities in their dynamics and processes of change (Boyd & Richerson, 1985; Cavalli-Sforza & Feldman, 1981; Mesoudi & Whiten, 2008).

An axiom of biological evolution is that it is not directed, meaning the sources that provide the raw material for natural selection are blind to their effects on the fitness of an organism. In other words, genes can only produce random variation, from which advantageous traits are selected and retained by way of the reproductive success of the organism carrying those genes. Thus, the fact that a given change would be advantageous does not increase the likelihood of its *appearance* and is not in itself a reason for a trait to arise. This critical idea is deceptive in its simplicity and can be difficult to apply, and as such is frequently violated in evolutionary theorizing. As Gould and Lewontin (1979) and later Gould (1989) have described, many proposed evolutionary scenarios fall prey to a tendency to equate current utility with reasons of

origin. That is, theories often carry an implicit assumption that evolution inevitably leads to progressively more complex and sophisticated traits and behaviors – in other words, that evolution is inherently *directed*. However, because the mechanisms of biological change are blind, progress is not guaranteed, and stasis is in fact the norm, particularly when a trait or strategy is especially successful (Gould, 1996).[1]

There are some reasons to caution making a direct analogy to cultural evolution concerning the idea of directed change. In contrast to biological evolution, variation in cultural evolution is generated by intentional beings with goals, needs and desires, and selection stems from the psychological and social features of cultural transmitters themselves. As such, it is possible to argue that innovation from the insight of those transmitters results in a more directed process, with more advantageous strategies or artifacts arising than sheer random variation would predict. However, the extent to which cultural evolution is directed in this sense is contested, as cultural change can often resemble the 'blind' processes of biological evolution (Mesoudi, Whiten & Laland, 2006). That is, improvements in cultural artifacts are frequently the result of copying errors or unintentional alterations that are only recognized as advancements after the fact and retained. This indicates that evolutionary explanations that invoke insight into cultural advancements may not be the most likely or viable solutions. In addition, studies on cultural processes involved in communication suggest that even non-directed cultural evolution can produce improved communication strategies. A number of computational models and behavioral experiments demonstrate how 'blind' cultural mechanisms involved in repeated transmission and learning can shape communicative systems to take on the appearance of 'design' for enhanced functionality (Batali, 1998; Kirby, 1999; Kirby, Cornish & Smith, 2008). In these studies, low-level forces have amplified and unanticipated effects as a result of the complexity that arises out of many communicative interactions taking place over an extended period of time. The emergent properties of such systems and processes are therefore a potential alternative source for more effective communicative strategies apart from intention-based forces.

[1] Note that this position is not equivalent to or an adoption of Gould's (1977) saltationist theory of evolution, which holds that evolutionary changes typically take place rapidly and in large steps. The point being emphasized here is related to larger patterns of change – how frequently and for what reasons changes happen – not the dynamics of how those changes take place.

2.4 Discussion

The concept of parsimony is an important tool for any inquiry regarding evolution, and particularly for one attempting to understand the processes that resulted in a complex behavior like human symbolic communication. This question has many interacting factors, and therefore many potential competing evolutionary explanations. With this principle in mind, however, we can rule out many hypothetical solutions and identify the most viable scenarios.

The embodiment perspective discussed above bears directly on the issue of continuity. The particular anatomical, cognitive and social features of pre-symbolic communicators will in part determine how evolutionary processes could have operated to result in the emergence of symbolic communication. Thus, it is critical in the present investigation to explore the communicative and evolutionary consequences of those features.

This discussion of evolutionary theory highlights three primary principles with which this thesis aims to be consistent: parsimony, continuity and the avoidance of appeals to intentional forces. These ideas will be used to evaluate existing theories and formulate new hypotheses in later Chapters.

3. Signs and symbols

This section will examine the nature of signs, or representations for communication, as they are understood in semiotics. The concepts and principles of sign theory provide a framework for understanding the defining characteristics of symbols and how symbolic signs could arise from non-symbolic signs through communicative processes. Different sign types will be described, and the key factors distinguishing symbols from other sign types will be identified. These differences will be discussed in terms of implications for communication and emergence.

Before discussing sign types, a brief terminological clarification is necessary. The term ‘symbol’ is used for a variety of purposes with many different, often conflicting, meanings. As it is used here, symbol refers to a sign – an intentional representation

for communication – with a particular form-meaning relationship (this relationship is described below) that is context-independent (can be displaced in time and space from its referent). This is a distinct use of the term from a number of other contexts. For example, one frequent meaning for *symbol* is 'functional reference' (Marler, Evans & Hauser, 1992). Functional reference is used to describe certain animal alarm call systems. However, such calls are not like the signs of interest here, as they are not used outside predator situations and therefore not context-independent. A different use of the term *symbol* is Deacon's (1997). He conceived a complex network of interrelated signs wherein symbols refer to other symbols and the relationships between them, as opposed to referents in the world. While one could argue that in order to exhibit the properties of symbols in language there must be some kind of association between symbols beyond associations for objects in the world. However, for present purposes, it is not necessary to specify whether or how symbols relate to each other in these ways. As this thesis is investigating the *emergence* of symbolic communication, it is not immediately relevant to account for or consider this more elaborated version of symbols; at present we are interested in individual symbols, not *systems* of symbols. Finally, another frequent use of the term 'symbol' comes from cognitive science and refers to any representation that is manipulated in a computational process (Fodor, 1975; Zenon & Pylyshyn, 1980). Symbols in this sense are not communicative signs – they are abstract units of mental content involved in subconscious cognitive operations. As the present discussion is concerned with outward, embodied representation for communication, this cognitive science notion of symbol is not directly relevant.

3.1 Signs and sign types

A sign, in the most basic sense, is a form that carries meaning. Peirce (1931-1958) first categorized signs into three types: icons, indices and symbols. These three types differ from each other in how a sign's form relates to its meaning. This section will describe what constitutes this relationship – a sign's *ground* in Peircian terms – for each category of signs. While the nature of signs and distinctions between sign types continues to be actively discussed in semiotics, some basic principles are well established and widely accepted. This review and analysis will take the position, also

proposed by Peirce, that sign relationships involve the interpretations of sign users, implications for which will be elaborated on further in the following section.

An icon is 'a sign that is made to resemble, simulate or reproduce its referent in some way' (Sebeok, 1994; pg. 10). Iconic signs represent a referent by virtue of some aspect of our sensory experience of that referent; they are the 'transformation of perception into representation' (Danesi, 1999; pg. 33). Iconicity, then, is representation based on similarities between features of a form and features of a meaning, which theoretically can occur in any of the visual, auditory, tactile and olfactory modalities. Icons are frequently used in human communication, from pictures in instruction manuals to gestural pantomiming in the game charades to onomatopoeia in speech.

Indexical signs are based on concrete physical, temporal or causal associations that occur in our experiences in the world. The presence of one component of an index evokes awareness of the other. For example, smoke is an index of fire, and thunder is an index of lightning. Note, however, that the relationship between smoke and fire is not an instance of an indexical relationship between a sign and referent *for communication*. The most common and prominent example of indexical representation in human communication is deictic pointing, which is based on physical location and division of space. The typical pointing gesture involves a single digit extended toward an intended referent. In this case, the direction of the digit indicates the direction of the referent in space and identifies it by guiding an interlocutor's gaze to its location. Pointing occurs when a referent is present, but it is also frequently used to represent an absent referent by indexing a previously occupied location or points in space assigned to different referents.

Signs that make use of indexical relationships are ubiquitous in human communication, although these are typically built upon iconicity. For example, interstate road signs in the United States use the image of a tent to indicate that campsites are present at a given exit. The image itself depicts a tent, but the sign refers to spaces where visitors can set up tents or campers. Thus, the sign represents campsites by resembling an object associated with campsites. In this way, communicative indices make use of associations in our experiences by way of forms

that resemble one component of the indexical relationship. This blending illustrates that signs may not be exclusively one type, a topic that will be addressed in the following section.

In contrast to icons and indices, the forms of which bear a direct relationship to their referents, the relationship between form and meaning for symbols is arbitrary. Symbolic forms are connected to meanings without regard to resemblance in features or concrete association. As a result, symbols represent their meanings through cultural convention, or a shared understanding between individuals that a given form stands for an agreed upon meaning. While icons and indices can be and often are conventional, symbols *must* be conventional in order to serve as effective communicative devices. That is, because the meaning of an arbitrary sign is not retrievable from features its form, successful communication via symbols requires that communicators possess the same knowledge that a form stands for a particular meaning.

With the different representational relationships of the three sign categories in mind, it becomes apparent that icons and indices together comprise a more general category of nonarbitrary representation. Nonarbitrary signs have a direct relationship between form and meaning – either through resemblance in form or concrete association. In contrast, arbitrary signs, or symbols, have an indirect relationship that exists as a social agreement between communicators.

3.1.1 Indices vs. symbols

A closer consideration of the distinction between indices and symbols reveals it is somewhat less clear than the description presented above. If an index is based on an experienced association between form and referent, it could be argued that signs otherwise generally agreed to be symbols are similarly characterized by reliable associations.

The standard distinction rests on symbols' referents not being *concretely* associated with their forms like indices are. This reasoning is based on the fact that symbols may never be used at the same time that their referent is physically present, as one can learn about, say, zebras from others' accounts and later talk about them without ever

having seen a zebra or even a picture of a zebra oneself. While symbols may not involve a physical correlation in this way, their forms are nevertheless associated with their referents; the two do reliably co-occur, in that a word is used when communication is about its referent. For example, the word 'cat' always occurs when the topic of discussion involves cats in some way, which is a kind of association between a form and a concept to which attention is directed. In this sense, the relationship is not arbitrary, because the word and the communication about the concept reliably co-occur. If we accept this as an association, then it becomes difficult to draw a clear line between indices and symbols.

One way to reconcile this problem is by appealing to an important difference in the correlation between symbols with their referents and indices with theirs. Although symbolic forms are associated with their referents, these co-occur only in the minds of communicators as they express and comprehend meaning conveyed through signs. In other words, the association for symbols exists internally to the semiotic process itself and nowhere else. Indexical relationships are exploited for communicative purposes only after they have been observed and reinforced in non-communicative experience in the world. The connection between symbolic forms and their referents does not require nor seem to involve a process of reinforcement and co-option, as they are inherently communicative and immediately understood as signs. The relationship between symbolic form and referent can be said to be arbitrary in a way that that of indices is not because the connection for symbols is interpretable only if we first assume that the form is being used as a sign. The connection is inferred based on the perceived intent of another person to communicate a meaning with a form, not directly through our own independent experience in other contexts. This is somewhat along the lines of Deacon's (1997) conception of symbols, under which the connection between a new symbol and referent is established by reference to other symbols. In this way, the distinction can be thought of as relying on direct, concrete association versus semiotic or indirect communicative association.

The blurring of indices and symbols is also apparent in child language development. It may be the case that in the early stages of children's acquisition of language their understanding of words is through indexical reference. Much of what is communicated by and to children involves objects and events in the immediate

environment (Lust & Foley, 2004), which means the words they hear are concretely associated with their referents. Eventually, however, children begin to learn words for and communicate about things beyond what is physically present. Moreover, they do so after a single exposure to a word, without the need for repeated observations. It is both intuitively appealing and theoretically useful to posit that at some point in this process the way children learn and interpret signs qualitatively shifts. This shift can be captured by characterizing their later use of signs as symbolic.

Although the distinction between indices and symbols may be more complicated and subtle than distinguishing either type from icons, it is theoretically useful to recognize the differences that do exist. Doing so helps to understand the unique properties of language and how linguistic signs could have developed from other types of communication.

This section has described sign categories in terms of forms and meanings. However, in order to understand specific, individual signs, an additional element concerning a sign's user(s) is required. How these different relationships constitute a given sign involves an act of interpretation by a user, which is addressed in the following section.

3.2 Sign interpretation

A central notion in semiotics is the idea that forms do not intrinsically carry any meaning. For a sign to represent something requires it to be *interpreted* as such (Sebeok, 1994; Eco, 1979). Not only is *that* a sign carries meaning interpretation dependent, but *how* it does so is as well. In this way, a sign's meaning and status as an icon, index or symbol critically depends not on form alone, but on how that form is understood *by users*. This issue involves nuances and questions that the discipline of semiotics continues to address; however, this thesis will take the *context-dependent* perspective of users as a foundational and guiding principle for conceptualizing symbols. The present inquiry is concerned with a very specific sign system – words understood to be arbitrarily related to their meanings – created and communicated by very specific users – embodied human beings. The psychological and behavioral profile of users will affect how signs are perceived and used, which is a determinant

of sign type. Thus, the interpretation of users in context is highly relevant from an embodiment perspective, and it is an essential tool for understanding how symbolic language emerged.

The interpretive element of signs is especially apparent with symbols, as they bear no direct relationship to their referents and have no connection to their meanings outside knowledge in the minds of users who possess a conventional system with which to interpret them. However, the same is nonetheless true for icons and indexes, as someone must *recognize* the resemblance or association between a sign and its referent. This section will describe the interpretive processes underlying each sign type and discuss the implications of sign interpretation for understanding how forms could transition between nonarbitrary and arbitrary signs.

For icons, interpretation is based on recognition of similarities between features of a sign's form and features of its referent. As a result of this inherently interpretive process, the number of iconic features appearing in a form functioning as an icon can vary greatly. A given referent will have many different features, and an iconic representation may depict a few or many of these. In many cases, fairly minimal and schematic iconic resemblance is sufficient for an iconic relationship to be perceived. Moreover, iconic representations can vary not just in *how many* features of a referent are included, but also *which* ones are depicted. The particular set of features depicted can itself be meaningful and reflects the communicative demands of a given situation. Thus, the possible ways an icon represents its referent through particular combinations of salient features varies enormously and will depend on a number of factors, including why the sign is produced, who is producing it, the medium through which it is produced and the wider context of sign production. It should also be noted that a prerequisite for understanding an iconic sign is familiarity with the referent itself. Regardless of how many iconic features are present in a form and how elaborately depicted a referent is, without knowledge of the referent, those features cannot be perceived.

For indices, interpretation is based on awareness of a correlation; a sign is indexical if an association is perceived between its form and referent. As with icons, one must

have gained from experience knowledge of the correlation between the components of an index in order to interpret one as standing for the other.

The critical role of interpretation is most obvious with symbols. Unlike icons and indices, interpretation of symbols relies on cultural knowledge of mapping arbitrary forms to meanings. Recognition in this sense comes from a socially shared awareness that a form stands for a meaning, regardless of how features of the form relate to features of the referent or whether the form and referent are reliably associated in experience. Note that implicit in this knowledge is the assumption that the relationship between form and meaning holds for others as well, as it is a representation *for communication*.

The differences between sign types in how forms relate to meanings, and by extension how they are interpreted, reveal another important distinction between symbols and nonarbitrary representations. In addition to the requirement, shared with icons and indices, that to communicate with a sign its form must be perceived as representing meaning, symbolic communication involves another underlying factor: conventionalization. Because the interpretation of a symbol is not knowable through any information contained in its form, effective communication via symbols demands that interlocutors share the same pre-established, arbitrary form-meaning mappings – *they share the same interpretation of signs.*

Conventionalization is a functional consequence of – and prerequisite for – using arbitrary signs as communicative devices. In this way, interpretation and conventionalization are not entirely independent. Interpretation is an individual-level phenomenon that feeds into conventionalization, a cultural-level phenomenon. Conventionalization of a symbol, thus conceived, is the cultural propagation of a certain interpretation of an arbitrary sign across a community of individuals. While interpretation and conventionalization will both be present in an established symbolic communication system, the two factors become disconnected from one another in the conditions of novel symbols. We can disregard as highly improbable that, without any previously established symbols, multiple people would spontaneously and simultaneously decide that the same form arbitrarily stands for the same meaning. As such, we can assume novel symbols initially will exist only in the mind of a single

individual (*how* such interpretations might come to exist will be discussed below and addressed in regard to the focus of this thesis in later chapters). Therefore, dissecting symbolic communication into the factors of interpretation and conventionalization becomes especially important when trying to understand the emergence of symbols. This issue will be discussed further in section 3.4 below.

Finally, while conventionalization is necessary for symbols, it is by no means exclusive to symbols. Nonarbitrary signs can and often are conventional, like the road sign example above. The conventional aspect of symbols is emphasized here because symbols are *obligatorily* conventional, while nonarbitrary signs are not. Thus, an explanation for symbols must account for this factor.

3.2.1 Sign type blending

A consequence worth noting of interpretation is that the categories of icon, index and symbol are not mutually exclusive (Keller, 1998). The case of modern Chinese characters demonstrates this point. Some characters originated as iconic pictorial signs, but their forms were altered substantially over time. Most current users are not aware of the signs' development and understand them to be arbitrary and symbolic. In contrast, others who have studied their histories know the original and intermediary forms and can therefore recognize surviving iconic features, when they exist, even if very little of the earlier form remains. In this way, a character may be iconic to one user while purely symbolic and arbitrary to another who does not share the relevant historical knowledge. This example illustrates how a single form's representational status can differ between users as a function of the different knowledge and experience those users possess. As a result, form alone cannot unambiguously indicate how a sign represents meaning.

In addition, the layering of iconicity and indexicality in signs discussed above (section 3.1) reveals that sign types are not mutually exclusive for an individual interpreting a single form. A sign may refer by way of both iconic and indexical features, making it neither purely an icon nor index, though wholly nonarbitrary. This intra-sign, intra-individual blending of sign type would seem to be possible only for nonarbitrary signs. If, under the perspective adopted here, the distinguishing characteristic of symbols is an arbitrary relationship between form and meaning, then a sign that is

symbolic cannot involve iconicity or indexicality. However, it is possible for individuals to hold idiosyncratic nonarbitrary perceptions of forms used as symbols by the majority of a communicating population while not assuming that others share that perception. For instance, if in learning a second language a person acquires a new word by perceiving it to 'sound like' its referent and uses that iconicity to remember the word's meaning, that person's interpretation of the sign relies on nonarbitrary properties. That same person also likely understands that others do not use the word in the same way, and in communicating with it this person assumes the word is functioning symbolically (though their understanding is unlikely explicit or thought of in these terms). Thus, a single form may be understood as both nonarbitrary and symbolic for an individual, but only as a result of a divergence between their subjective interpretation and others' assumed interpretations of the sign. The blending of nonarbitrary and symbolic sign type, then, is unlike that of nonarbitrary blends. While the latter exists in a single interpretation of a form, the former makes reference to two separate interpretations of the same form.

The possibility for sign type blending reveals that sign categories are ideal types of signs in theory, but in practice many *actual* signs embody a combination of representational qualities – an issue considered in Peirce's (1931-1958) work. In this way, a sign that is considered symbolic may in fact not have a fully arbitrary interpretation by all users. However, if that sign is used and understood by virtue of convention and not by means of any *shared* nonarbitrary perceptions in its form, this indicates the sign can be considered a symbol. In this way, a sign's interpretations across individual users may not be consistently or purely arbitrary, but the lack of consistent and robust nonarbitrary interpretations to enable comprehension without convention renders it effectively arbitrary. It is thus in this necessary dependence on convention sense that we can conceive of symbols as 'arbitrary'.

One final note worth addressing on interpretation and sign types is Sonesson's (1997) proposed distinction between primary and secondary iconicity. In these terms, primary iconicity is when a form is understood to be an iconic sign because it is perceived to resemble a particular meaning. Secondary iconicity is when a form is perceived to resemble a particular meaning after it is understood to be a sign by way of cultural learning. In other words, some signs may be perceived as iconic only after

the connection between form and meaning is first established through convention. Note that the realization of secondary iconicity is only a possibility. That is, it is not guaranteed that or necessary for the iconic relationship to be perceived, as the sign can be understood through convention alone and still used effectively in communication. In cases of secondary iconicity, then, it may be that some users perceive a sign as iconic while others do so as symbolic. The implications of this will be discussed further in section 3.2.3 below.

3.2.2 Context-specificity

The interpretive nature of signs entails both that i) any form – that is, any entity or object in the world in any modality – is potentially a sign, and ii) a single form can potentially function as many different signs of each type depending on how perceivers interpret its connection to particular meanings. The factor that distinguishes non-representational objects – be they sounds, images, whole objects or any other form – from signs is the fact that signs are forms that are interpreted as meaningful by certain individuals. Similarly, the factor that distinguishes whether a given form is functioning as an icon, index or symbol is how the form is perceived to relate to its meaning by certain individuals. How a form is perceived by particular communicators using a sign in a particular context is a function of those users' shared knowledge and experience – related to the concepts of common ground (H.H. Clark, 1996) and relevance theory (Sperber & Wilson, 1987). Sign users will share a physical, social, cultural and/or interactional history in which the sign's form-meaning relationship has been established, which bestows them with a shared awareness of that relationship and allows them to communicate successfully with the sign. As such, separate contexts with users who have correspondingly different knowledge and experience can yield different sign relationships for the same form. Distinct contexts resulting in unique and independent sign interpretations can range from a single, temporally bound conversation between two people to an entire culture across centuries.

An illustration will help to demonstrate that a sign's status as an icon, index or symbol critically depends on how it is perceived by particular users. As an example of how one form can vary between sign types and meanings depending on context, consider the image in Figure 1.1. This form could iconically depict a number of referents, but

for present purposes we can pick an immediately obvious interpretation and imagine that it might be used to represent cats' whiskers. In this case, it would be an iconic sign, as features of the form directly depict features of the referent. In addition, as it may be evocative of cats in general for most people familiar with feline creatures, we can imagine that it could be used to represent something associated with cats, like a litter box. In this context, the sign would be an index (by way of iconicity for the associated concept); the form would be perceived as resembling whiskers, which are a part of cats, which use litter boxes. Furthermore, it is conceivable that, perhaps in the interest of secrecy, the image could be used as a symbol by an intelligence organization to represent something arbitrarily, like the now-convicted underground international nuclear arms trader Viktor Bout.[2] The context-internal, unique histories and awareness shared by users in each of these hypothetical situations is what allows for the same form to function as multiple signs and be interpreted in vastly different ways.

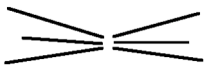


Figure 1.1. A form that in the hypothetical situations described functions independently as an icon for whiskers, an index for litter box or a symbol for Viktor Bout depending upon its users and context.

In the first case, in which the sign is iconic, it is recognized as resembling the referent itself, which is possible due to knowledge and experience of cats and their whiskers. In the case of the index, the resemblance in form is also recognized and similarly dependent on knowledge of cats, but a further step is made to an associated concept, which requires additional knowledge about cats and their use of litter boxes. Finally, the symbolic sign is understood not by any recognition of corresponding features between form and referent or by experiencing their co-occurrence, but only by users' shared knowledge that the image stands for 'Viktor Bout' in the community of their organization. The crucial point revealed by these hypothetical situations is that the

[2] I have made this third hypothetical situation involve the intentional creation of an arbitrary sign for illustrative purposes; however, evidence discussed later in this thesis indicates deliberate invention of symbols such as this is atypical.

key determinant of sign type in each scenario is the interpretation of the sign's users. The referent changes while the form remains constant, and this change in referent is due to different users communicating with a sign for different purposes. Moreover, the change in referent corresponds to a change in how form relates to meaning, which also resides in the perceptions of sign users. This demonstrates the fact that a sign is not merely a form connected to a meaning, but a form connected to a meaning *by a perceiver* (Sebeok, 1994). A sign, therefore, has three components: a form, a referent and a user.[3] It follows that the meaning and semiotic status of a given sign cannot be identified without reference to the user, as the relationship between form and meaning that determines that status exists in the user's interpretation of the sign.

It is important to note that it is not possible to know with certainty how signs are interpreted by users in the contexts described above by looking at the form and the referent in each case and trying to discern objectively the connection between the two. In the iconic and indexical case, seeming resemblance between form and referent or an associated concept is suggestive evidence that the form is nonarbitrary, though it is possible that the same form could be used to represent the same referent in a different context by others who, for a number of possible reasons, do not recognize the resemblance and/or correlation. Thus, this type of evidence is insufficient to reliably determine signs are icons and indices. In the case of the symbol, objectively establishing that the form arbitrarily represents its referent becomes even more difficult. For instance, with knowledge of the referent, we might suspect, and cannot rule out, that the form is meant to depict Bout's characteristic mustache. Or it could be intended to resemble the explosion of a bomb, a concept associated with Bout's business dealings. In the case of all sign types, then, without knowing how a sign's users perceive its form, we cannot determine with certainty if they are using it as an icon, index or symbol.

One final point on an objective approach to discerning sign type is that doing so becomes more difficult the simpler and more abstract a form is. Somewhat

[3] The *ground* mentioned previously is also sometimes included as part of a sign (for example in Ahlner & Zlatev, 2010). As the connection between form and meaning ultimately resides in the user, including the interpreter is sufficient for present purposes.

counterintuitively, the more schematic a form, the more potentially iconic it is. This is especially true for distinguishing icons from arbitrary symbols, as humans perceive resemblance in instances where only slight similarities exist, as evidenced by the common activity of looking for shapes of objects in clouds or the tendency to see images of religious figures in various artifacts that were not intentionally designed to depict specific images (for example, the burn patterns in a piece of toast). When a form contains one or only a few features, those isolated features can be mapped onto similar features of many different and widely varying referents, which may not on the surface be perceived as resembling each other but happen to have that one particular feature in common. The absence of additional features means there is nothing to conflict with potential referents' many other features. Due to this high match-ability, together with humans' tendency to seek out and perceive iconicity, it is extremely difficult to rule out that a sign with a minimal form is not interpreted as iconic in some way by some of its users.

3.2.3 Sign type transition

An additional consequence of interpretive processes defining signs is that, at least as observed in modern human communication, changes in sign type for a given form can occur simply through the use of a sign over time (Keller, 1998). In the same way that a form can be iconic for one individual and symbolic for another in independent contexts, this asymmetry can also arise within a communicative exchange when one individual does not perceive the intended iconicity of a sign and interprets the form as arbitrary and symbolic. If and how semiotic transitions occur will depend on the perceptual and social features of sign users, the number of dimensions in which forms resemble their meanings (when nonarbitrary), the number of channels through which forms are expressed, and the medium(s) in which forms are expressed, among other factors, and therefore should be considered in relation to the specific conditions of a given communicative context. The process by which imperfect communication through nonarbitrary signs could result in symbols will be examined in more detail in Chapters 2 and 3. For now, the point of interest is that, for purposes of investigating the emergence of symbols from nonarbitrary signs, it is useful to conceive of symbols in the negative: Interpreting a form as symbolic can simply involve *not perceiving* iconic or indexical features. Taking this perspective is key to understanding research that will be presented in Chapter 2 and the arguments put forward later in this thesis.

The next section will briefly discuss implications and other issues related to different sign types.

3.3 Comparison and implications

In what ways do icons and indices differ from symbols beyond those related to interpretations of signs described above? The different relationships between form and meaning also have communicative and cognitive implications. This section will review these implications and discuss how they relate to the question of the emergence of symbolic communication.

3.3.1 Nonarbitrariness

Nonarbitrary representations have communicative advantages by virtue of the direct relationship between form and meaning. The perception of iconic and indexical features is based on knowledge that naturally arises out of our experience observing and acting in the world, which means that knowledge will be largely shared across individuals simply as a consequence of encountering similar objects and events. As a result, icons and indices can, at least potentially, be interpreted directly through this common sensorimotor and associational knowledge without the need for an agreed upon and learned conventional system. As symbols are arbitrary, they are disadvantaged in this sense. Because specific cultural knowledge must be acquired in order to understand symbolic signs, they cannot be used to communicate unless both parties share that knowledge. This point is exemplified by the fact that speakers of different languages who cannot understand each other's speech can nevertheless communicate, to some degree, through pantomime.

Another advantage of having an immediate connection to our sensory systems appears to be an increase in memory for communicated information when that communication includes nonarbitrary representations. When speech is accompanied by iconic and indexical gestures – as frequently occurs in natural communication – more information from the overall message communicated by the two modalities is remembered (Beattie & Shovelton, 1999; Feyerseisen, 2006; Kelly, Barr, Church & Lynch, 1999; Kelly, McDevitt & Esch, 2009; the multimodal nature of language will

be explored in depth in Chapter 3). It may be that because nonarbitrary representations like gestures directly engage our sensorimotor knowledge and capture rich aspects of our experience, they provide an evocative mode of expression that is especially powerful and perceptually salient. In this way, nonarbitrary signs seem to be expressive in a manner that is particularly compelling to our senses.

3.3.2 Arbitrariness

While the arbitrariness of symbols may not have the same expressive and mnemonic effects, it does appear to have benefits of its own. In a computational modeling study, Gasser (2004) demonstrates that arbitrary signals are a superior signaling strategy when communication involves many and related meanings. Assuming some pressure for economy in signaling, iconic forms will be limited in the number of features through which referents can be portrayed. Given this constraint, a single form could resemble two or more very similar meanings, but which one of these the sign is intended to represent could not be distinguished based on form alone. Thus, if the meaning space of a communicative system is densely covered – if the meanings conveyed share many features – iconic representations will become increasingly ambiguous. Because arbitrary signals are disconnected from their meanings, they are free to take a wider range of forms. This freedom allows them to remain distinct from one another even as the number and similarity of meanings increases, and thus avoid the ambiguity of nonarbitrary signs. As a result, arbitrary symbols offer the potential for infinitely precise and expressive communication.

Another feature of arbitrary symbols is their detachment from immediate perceptual experience, and there may be some advantage to limiting the interaction between a sign's form and the sensorimotor system. According to A. Clark (2008), symbols serve as powerful boosters of cognitive potential as a result of this disconnect. Symbols function as simplified stand-ins for things in the world; they are abstractions of what is referred to that are not bound to any particular context or perceptual event. Labels, as Clark calls them, therefore become separate objects of thought that allow for new cognitive possibilities – a view that is mirrored in Tomasello (1999) and Deacon's (1997) conception of symbols and their function in language and thought. By condensing a rich and varied aspect of experience into a single representation, symbols reduce the computational load of many cognitive processes. Thinking about

relations between complex features of the world is thus converted from a complicated problem to a much simpler one. This allows for more and more regularities in the environment to be recognized and labeled with ever-increasing levels of abstraction. Once established, symbols can then regulate and structure cognitive processes, aiding selective attention and guiding reasoning.

This claim about symbols seems to gain support from empirical evidence. Chimpanzees were able to solve problems concerning second order relations only after learning labels for the first order relations (Thompson, Oden & Boysen, 1997). Without labels, the chimpanzees were able to indicate correctly if two items were the same or different but were unable to do so when the task involved extending this concept to a comparison of two *sets* of items; as in, if the 'sameness' of one set is shared by the other set. Once the chimpanzees were taught labels for 'same' and 'different' that they could assign to individual sets of items, however, they could then respond correctly to comparisons of sets. As additional evidence, Smith & Gasser (2005) survey research on human ontogeny and argue that encountering a symbol system like language profoundly alters cognitive development. The order of development – beginning with grounded, multimodal learning of how to move and act in the world and progressing through increasingly higher level cognitive milestones – suggests that acquiring language enables children to develop more complex and sophisticated modes of thinking and acting by providing simplified, arbitrary representations of rich and perceptually-loaded concepts.

3.4 A paradox in inception

A problem seems to arise when we try to apply the characterization of symbolic signs as presented – arbitrary and conventional – to the emergence of symbols. If we are thinking about symbols arising from non-symbolic communication, it necessarily follows that there must have been a time at which non-conventionalized arbitrary signs were used in communication. This may at first appear nonsensical: If no shared system of arbitrary signal-meaning mapping existed, why would someone produce a sign that others would not understand? One possible solution is that an individual or individuals would create symbols and teach others how the sign should be used. Due

to reasons addressed in section 2 above and expanded on in section 3.5 below, however, an alternative to deliberate instruction is desirable.

The apparent paradox - symbols as obligatorily conventional versus the prerequisite of non-conventional signs being used 'symbolically' for shared systems to emerge in the first place – can be resolved by breaking up symbolic representation along the lines of the two underlying mechanisms identified in section 3.2: interpretation and conventionalization. Under this conception, interpretation of a sign as symbolic is possible without that interpretation being shared by any other individuals.

To understand how this works, note what is typically implied when a sign is interpreted as symbolic. Part of this interpretation is an understanding – or belief – that others attach the same meaning to a given signal, what I will call *virtual conventionalization*. Virtual conventionalization is what could induce an individual to use a sign to communicate with others who in actuality may not share their symbolic interpretation of that sign. In doing so, however, such an exchange could result in the interlocutor coming to adopt that interpretation, and therefore spur conventionalization. The relationship between interpretation and conventionalization in the emergence of symbolic communication will be examined in greater detail in Chapters 2 and 3. For now, it is sufficient to note that teasing the two apart allows us to understand how, in a stage prior to the emergence of shared symbolic systems, signs that do not meet the conventionalization criterion for symbols could nevertheless be used symbolically by individuals who interpreted them as such.

3.5 Summary and discussion

Icons, indices and symbols are similar in that they are representations for communication, but they differ in terms of form-meaning relationships. Icons and indices embody a direct relationship between form and meaning and can therefore, at least potentially, be interpreted without additional, communication-specific cultural knowledge.[4] In contrast, symbols have an arbitrary relationship between form and

[4] As recognition of icons and indices is also dependent on experience in the world, cultural knowledge regarding artifacts and other relevant aspects of a group's shared environment is part of this experience and will similarly underlie the recognition and

meaning, and by extension require users to possess shared knowledge of particular arbitrary form-meaning mappings in order to function as effective communicative devices. These two features correspond to the factors identified above as underlying symbolic communication – symbolic interpretations of forms and conventionalization of those interpretations. Thinking about the defining features of symbols in terms of sign interpretation provides an essential framework for explaining how and why symbolic signs may arise in a given context.

The inherent involvement of interpretation in sign use means a relationship between form and meaning, and therefore a sign's status as iconic, indexical or symbolic, is a subjective matter and implies that it must be discerned in relation to a specific context. In addition, it further implies the interpretation of users is integral to defining the relationship between form and meaning embodied in a sign, which was shown not to be retrievable from knowledge of the form and referent alone (section 3.2.2). In this way, a sign is constituted by an irreducible triadic relationship between form, meaning and user (Sebeok, 1994); without reference to the user, we cannot know how form relates to meaning. Thus, the perceptions of sign users in context will be considered the critical indicator of sign status.

The position taken here may seem somewhat rigid in regard to distinctions between sign types. Often iconicity is spoken of in terms of a gradient, as in a form being more iconic or more arbitrary (for example, in Garrod, Fay, Lee & Oberlander, 2007), which is meant to reflect differing levels of elaborateness or detail depicted in forms representing the same meaning. This is not entirely incompatible with an interpretation-in-context perspective, though it requires conceiving of the gradient as deriving from interpretations in aggregate. A sign might be perceived as nonarbitrary by some and arbitrary by others while all are unaware of these conflicting interpretations but are nevertheless able to communicate with the sign. This is likely the case for signs that exhibit secondary iconicity (Sonesson 1997; section 3.2.1), as their conventionality ensures mutual understanding even in the absence of shared sign type recognition. Thus, we could describe a given sign as 'more symbolic' if more

---

interpretation of nonarbitrary signs related to this knowledge. In contrast, the cultural knowledge of form-meaning mappings required for symbols is explicitly and specifically for communicative purposes.

people, on average, do not recognize any nonarbitrary features and rely on cultural knowledge to make the connection between form and meaning. In capturing the notion of symbolic communication as we generally understand it in regard to language, it is also useful to take a population-level perspective. Given that arbitrariness has traditionally been held as a fundamental feature of language (Hocket, 1960; Saussure, 1916), it is likely that most words are in fact perceived as arbitrary by a large portion of a population of speakers. In addition, although it is possible that some individuals perceive certain nonarbitrary features in a given word, it is unlikely nonarbitrary interpretations are the same *across* individuals. Moreover, these individuals probably do not believe or expect others to share their idiosyncratic interpretations. We can coherently speak of language as symbolic, then, because i) on average, most people perceive most words as arbitrary signs, and ii) individuals who might perceive nonarbitrariness in a given word do not rely on this when using it to communicate with others. The context-dependent interpretation of individuals becomes crucial, however, when symbolic conventions are not established in a population and communication takes place through nonarbitrary signs. In this situation, individual symbolic interpretations are what will feed the development of population-wide symbolic conventions.

Regarding the comparison of sign types, while nonarbitrary signs are beneficial in not requiring conventionalization, symbols vastly expand the communicative and cognitive potential of their users. As theoretical points, the advantages associated with symbols are clear. These factors are a major part of what gives language the capacity to communicate complex ideas with precision and clarity. However, the subject of this thesis is the emergence of symbolic communication in human evolution. The context in which this took place – embodied experience – involves a number of various forces in operation, many of which would act to increase, subvert or balance the actions of others. Thus, we should withhold drawing any direct conclusions based only on the advantages and disadvantages of symbolic versus nonarbitrary representations in isolation from other factors.

The issue of 'blindness' and intentionality discussed in terms of evolution in section 2 above is especially relevant to the question at hand. Given these considerations, it does not necessarily follow that any advantages for symbols would be discerned and

deliberately exploited by pre-symbolic communicators. First, without prior conventions, symbolic communication is not an immediately obvious or advantageous strategy (Deacon, 1997), as it requires multiple individuals to share the same arbitrary form-meaning mappings. As such, the advantages of symbols may only become apparent *ex post facto*, after a conventional system is learned and used. In addition, a striking historical example of another symbolic revolution in human cultural evolution illustrates how even modern humans are not inclined to have insight into a new representational strategy despite its ready availability and numerous advantages. While we do not know precisely when modern spoken, symbolic language first appeared, it is generally believed to roughly coincide with the emergence of anatomically modern humans, somewhere between about 100,000 to 200,000 years ago. Written language, however, did not appear until much more recently in human history, probably no earlier than about 5,000 years ago (Schmandt-Besserate, 1996). The advantages of a permanent, written linguistic system are quite obvious in the abstract, and this form of representation ultimately enabled the elaboration of the tools, social institutions and practices upon which our modern cultural institutions are crucially dependent. We can see, in retrospect, that human society prior to the development of a graphical symbolic representation of language would have similarly benefited from the adoption of a writing system. Pre-literate societies often have elaborate and rich cultural traditions and information that must be passed down across generations through extensive teaching and the memorization of lengthy oral 'texts', such as poems and stories (Ong, 1982). The time and energy invested in preserving these cultural traditions could be greatly reduced by the creation of permanent records. In addition, in the same way that spoken linguistic symbols reduce the cognitive demands of processing conceptual information (section 3.3), written linguistic symbols reduce memory loads and create opportunities for contemplating and manipulating conceptual knowledge, which aids and enhances thinking about complex problems (Olson, 1996; A. Clark, 2008). Despite these profound advantages, and despite the fact that symbolic representation already existed as an established means of spoken communication, writing systems were only developed in isolated parts of the world in relatively very recent history. Moreover, symbolic writing systems only came about after a long period of cultural evolution from what appear to be initially nonarbitrary visual representations (Olson, 1996). There appears to have been no 'insight' for the potential to represent language through graphical

symbols, but rather, the transition occurred as a result of imperfect transmission and learning across generations and cultures, leading to the reinterpretation of originally non-symbolic signs (Olson, 1996). The fact that modern humans – with a remarkable ability for innovation – have not universally discovered and invented visual symbols for linguistic information, and that the limited cases where this has happened represent a considerably late fraction of human history, suggests the likelihood of pre-symbolic communicators deliberately inventing symbolic representations is highly improbable. Thus, appeals to intentional construction should be avoided in an explanation for symbolic emergence.

The research on cultural evolutionary processes described in section 2 shows that alternative avenues may be available by which symbolic communication could arise. Complex systems, like communicating populations, have emergent properties, and these studies demonstrate that functional communicative strategies can take shape under certain conditions. In this way, cultural processes that do not involve the conscious, deliberate action of communicators may be capable of spurring the development of symbols. Because this alternative has explanatory potential, an investigation of the emergence of symbols should first look to such mechanisms before appealing to intention or insight.

Finally, with the ideas presented in this section in mind, it is possible to more precisely specify the symbol aspect of the question at the heart of this thesis. The issue of symbols is one of how symbolic *interpretations* of signs arise, and following that, how those interpretations *conventionalize* – how they are transmitted to and adopted by others. The core of the symbolic relationship – the element that connects form and meaning – resides in users' perceptions of a sign. Understanding where symbols come from therefore first requires explaining how forms come to be perceived as arbitrarily representing meaning, with a consideration of how interpretive processes establish and alter signs' representational status. In addition, it also requires explaining how novel symbolic interpretations of signs then become shared by others and thereby become functional communicative devices. These points are central to this thesis and will be discussed in relation to research presented throughout the following chapters.

The preceding three sections have discussed how the present investigation will be framed. The following section will review the concept of bodily mimesis, which provides a starting point from which this thesis will approach the emergence of symbolic communication.

4. Bodily mimesis

Donald (1991) proposed the bodily mimesis hypothesis in an aim to bridge the gap between the communication observed in extant apes and modern human language – namely, symbolic communication. Donald argued that because the gap is so substantial, an intermediate stage must have existed prior to and somehow enabled the development of modern language. One of the defining features of language is the use of conventional, arbitrary words, but this phenomenon is ultimately built upon the ability and inclination to produce intentional, referential representations for communicative purposes. This more basic capacity, manifested in the form of bodily mimesis, is what is claimed to have preceded the development of symbols and fully-fledged language. This section will review what bodily mimesis is and supporting arguments.

4.1 What is bodily mimesis?

After it was first proposed by Donald (1991), the concept of bodily mimesis was refined and expanded by Zlatev and colleagues (Zlatev et al., 2005). In the most basic sense, it is using the body to communicate with nonarbitrary signs. Bodily mimesis therefore involves crossmodality, volition, representation and a communicative sign function (Zlatev, 2008). In other words, mimesis is the intentional production of sensorimotor-based, referential signs in order to communicate some concept to an addressee. Creating and comprehending such representations relies on the ability to map one's own implicit sensorimotor knowledge onto some other modality, and vice versa. For example, translating the visual experience of another person's gestural movements onto your own proprioceptive system and understanding it as a meaningful representation. Simply put, bodily mimesis can be thought of as indexical and iconic communication, such as pointing and pantomiming, though it can also

apply to other modalities like auditory vocalizations. In this way, mimetic representations are based on sensorimotor experiences of objects, entities, actions and events in the world that arise out of embodied, lived experience.

Zlatev and colleagues (Zlatev, 2007, 2008; Zlatev et al., 2005) make a distinction between bodily mimesis and what they call *post-mimesis 1* and *post-mimesis 2*, which refer to the specifically linguistic characteristics of conventional symbolic reference and syntactic structure, respectively. Mimesis, post-mimesis 1 and post-mimesis 2 form a hierarchy of stages in which each successive stage includes features from and builds upon the previous one. Progression from mimesis to language is therefore not through replacement, but instead the overlaying of additional features. Symbolic representation can be thought of as a coding system for an underlying conceptual structure that can also be directly communicated through bodily representations. This thesis is focused on the transition from bodily mimesis to post-mimesis 1 – communication through symbols. Bodily mimesis would have served as a means to establish communication in the absence of such a coding system, which would have laid the foundation on which shared symbols could emerge.

4.2 Support for the bodily mimesis hypothesis

Donald (1991) and Zlatev (2008) present a large collection of evidence in support of the bodily mimesis hypothesis. That evidence and their arguments will not be repeated in full here, but some key points will be reviewed that are relevant to issues discussed later in this thesis.

A major line of evidence these authors point to is the human ability to communicate through nonarbitrary gestures and the fact that gestures of this type are produced together with and closely related to speech (the relationship between speech and gesture will be addressed in detail in Chapter 3). Zlatev (2008) notes that because symbolic language is vastly superior in conveying complex meaning, there is no reasonable explanation for this ability and propensity for mimetic bodily communication unless communication relied on such mimetic skills prior to language. He goes on to argue that the fact that communication today makes use of nonarbitrary bodily representations in addition to the powerful symbols of speech suggests

linguistic capacity is an extension of the ability to express meaning through sensorimotor knowledge. Donald (1991) notes that language and mimetic gestures are dissociated, in that aphasics who do not also suffer from apraxia (motor impairment) typically retain the ability to communicate through nonarbitrary gestures, a phenomenon that has been documented in more recent studies (Fex & Mannson, 1998; Kemmerer, Chandrasekaran & Tranel, 2007). These cases demonstrate that language encompasses and relies upon abilities underpinning mimetic communication, and that mimetic abilities do not require language. Donald (1991) claims this dissociation indicates that mimesis is a more basic ability upon which symbolic language has been built.

Other supporting evidence comes from research on language and embodiment. A number of behavioral studies demonstrate that language processing modulates the sensorimotor system (Boulenger, Roy, Paulignan, Deprez, Jeannerod & Nazir, 2006; Glenberg & Kaschak, 2002; Parrill, Bullen & Hoburg, 2010; Sato, Mengarelli, Riggio & Gallese, 2008), and also that motor actions affect language processing (Rueschemeyer, Lindemann, van Rooij, van Dam & Bekkering, 2010). Neuroscientific studies have shown that language processing activates body-specific areas of the motor cortex related to the meaning conveyed in that language (Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006; Hauk, Johnsrude & Pulvermüller, 2004; Masson, Bub & Newton-Taylor, 2008; Pulvermüller, Shtyrov & Ilmoniemi, 2005; Tettamanti, Buccino, Saccuman, Gallese, Danna, Scifo, Fazio, Rizzolatti, Cappa & Perani, 2005; Willems, Hagoort & Casasanto, 2010). These findings indicate bodily knowledge and representations provide structure and content to linguistic meaning, and Zlatev (2007) argues this is precisely what would be predicted from the hypothesis that language arose out of sensorimotor-based mimetic communication.

Finally, Zlatev et al. (2005) discuss primate studies, noting that apes engage in imitation similar to that of humans, and imitative abilities ultimately underlie mimetic communication. In addition, when enculturated in a human environment, apes can come to understand the communicative sign function. The authors argue that the fact that these capacities are present in apes suggests the ability to communicate through bodily mimesis was a plausible and probable evolutionary route in the human lineage.

4.3 Discussion

The bodily mimesis hypothesis holds that prior to the development of modern language humans communicated through nonarbitrary, body-based signs. The capacity for bodily mimesis enables intentional, referential communication without the need for a conventional symbolic code. Once this kind of communication becomes possible, it provides an infrastructure from which a symbolic system could potentially be constructed.

It is useful to make explicit an underlying assumption concerning the context in which mimetic, and later symbolic, communication would be used. This kind of communication rests upon relatively sophisticated intention-reading abilities and an inclination for cooperativeness (Tomasello, 1999; Tomasello, 2008). A sufficient level of these intersubjective qualities is an assumption of a theory that explains a transition from one type of intentional, referential communication to another. Although it is highly relevant and important to understand how the distinctly human trait of shared intentionality evolved, this thesis will not attempt to provide an explanation for this foundational feature of human communication and cognition. The arguments presented here take as a starting point a behavioral environment of skilled intention reading and cooperative activity, as described in Tomasello (2008).

Together with these intersubjectivity skills, the fact that we share an embodied existence with others means we can understand them as similarly acting and sensing beings. Meaning comes from knowledge of lived experience, and human communication results from the capacity and inclination to represent that knowledge for another person. One strategy of sharing meaning in this way is through the use of our bodies as iconic and indexical representational devices, or mimesis. Interpreting others' bodily movements involves the application of our own sensorimotor knowledge and an understanding of intentions. Because this strategy, as opposed to symbols, is direct, immediate and requires no cultural code, it is plausible to suppose mimesis is a more basic, phylogenetically older form of communication than symbols. Furthermore, that our bodies form the basis for linguistic representations today suggests that communicative sensorimotor representations played a similar role in the course of evolution, predating and providing a foundation for symbolic language.

The connections to primate studies highlight important distinctions between ape and human communication that will be useful in evaluating evidence in later chapters. Apes use gestures to communicate, but the referential status of those gestures – or having the communicative sign function (Zlatev, 2007) – is not firmly established.[5] For example, the same gesture is often used to accomplish multiple and very different communicative ends (Cartmill & Byrne, 2010), which suggests these do not represent meaning the way humans use signs to refer to stable concepts across contexts. In addition, gestures are often created and developed between two individuals who do not then go on to use that gesture to communicate with others. The prototypical example of this dyad-internal gesture development is what is called ontogenetic ritualization. Ontogenetic ritualization is the process by which two individuals repeatedly interact under similar circumstances and shape each other's behaviors (Tomasello, 1997). The 'nurse poke' used by infant chimpanzees to initiate breast-feeding is a gesture derived through ontogenetic ritualization. The process starts with the infant touching and moving its mother's arm to reach her breast. Over time, the infant has to execute less and less of this action series for its mother to recognize its intent and begin breastfeeding. Eventually, the sequence becomes so reduced that simply poking the arm is sufficient to initiate the activity. This 'poke' is considered a gesture between mother and infant, one that has become shorter in duration and simpler over time.

Notably, these gestures differ from human communicative gestures like bodily mimesis in two major respects. First, they are not used outside the limited circumstances of breastfeeding; that is, apes do not produce these gestures to communicate about feeding in other settings for other purposes. In contrast, human communicative gestures are not tied to a specific activity and are used to discuss the meanings they refer to or related topics independent of context. Second, such ape gestures are only used between the members of the dyad, which indicates the gesture

[5] Language enculturated apes may be an exception, as they can acquire language-like labels and communicate with them even when their referents are not present (Savage-Rumbaugh, Shanker & Taylor, 1988). This latent capacity is important for finding continuity between modern humans and their supposed predecessors, but the nature of ape gesture without atypical intervention is relevant to understanding processes of sign development that will be addressed in later chapters.

is not understood by the apes as a form that carries meaningful content that can be conveyed to others. In contrast, human gestures are produced to *represent* a meaning that can be interpreted by different individuals across different contexts. In this sense, ape gestures appear more akin to the real actions themselves, abbreviated in form though they may be, and therefore lack the communicative sign function. Indeed, ritualization of this sort can be seen in human interactions, and similar reservations hold. For example, what starts as a child reaching for an adult's arms in order to be picked up can eventually become a simple raising of the arms, which is sufficient to initiate the picking up process by the adult. Would we consider this movement a symbol, let alone a sign of the kind under investigation here? Neither children nor the adults involved in this act use such gestures for communicative purposes with others or outside the act itself, again suggesting the movements are not representational in the sense of having the communicative sign function. The fact that the process of ritualization does not appear to result in signs even for modern human communicators suggests this process may be of limited relevance to the question of symbol emergence.

Finally, some researchers have characterized ritualization-derived ape gestures as iconic (Tanner & Byrne, 1996; Hopkins & Savage-Rumbaugh, 1991). However, Tomsello (2008) argues this iconicity only exists for human observers who recognize it as such; they are not iconic *to* the apes. At the end of the ritualization process gestures may appear iconic because they are rooted in purposeful actions, but as they lack the communicative sign function for the reasons above, it is doubtful they are interpreted as nonarbitrary signs by the apes. These and the other differences from bodily mimesis described above will be useful when examining processes of sign creation and development discussed in later chapters.

A final point to address is the placement of bodily mimesis on the human evolutionary timeline. Proponents of the hypothesis do not include a precise point at which bodily mimesis was established in the human lineage, though it is thought to have arisen mostly likely sometime after the appearance of *Homo erectus* and more advanced tool manufacturing and material culture, as these would presumably require the more human-like imitative skills that underlie mimesis. Although pinpointing when in human evolution such abilities might have appeared is an important research question,

this thesis will not attempt to address chronological issues directly. Instead, it will take a relative perspective and focus on what would have followed the development of bodily mimesis, though the issue of the pace of a transition to symbolic communication will be discussed briefly in Chapter 3.

Separating symbolic communication from the more general and basic communicative sign function as proponents of bodily mimesis do (Donald, 1991; Zlatev et al., 2005; Zlatev, 2007, 2008) is an important and useful theoretical distinction. In this view, symbols are not an entirely separate phenomenon, but are instead built upon the more basic capacity for intentional representation. Donald, Zlatev and colleagues have offered a thorough account of how bodily mimesis may have arisen in human evolution prior to symbolic language. A theory of the emergence of symbols, then, can take as its starting point this representational capacity. This thesis will investigate that emergence and will take as a premise the pre-existence of mimetic communication.

5. Overview

The preceding sections have identified the premises and framework under which this investigation will proceed. Embodiment theory places the question of symbols in a context with particular communicators that possess certain causally relevant modes of expression and perception. Evolutionary principles can constrain and guide the formulation of an evolutionary scenario involving such embodied beings. A consideration of semiotic theory reveals the explanatory factors that must be accounted for in a theory of symbol emergence. Finally, taking bodily mimesis as a starting point, the question of interest here can be further specified to ask how conventional symbols developed out of mimetic – or nonarbitrary, body-based - communication.

With this framework established, it is now possible to review and evaluate evidence in order to undertake the present investigation. The following sections will briefly outline the research examined and carried out in the remaining chapters of this thesis.

5.1 Chapter 2 - Emerging sign systems

The emergence of symbols in human evolution is one instance of the initiation and evolution of a sign system. A number of other examples of changing communication systems have been documented and analyzed. This research can reveal if and what general features and processes characterize sign systems, and thereby inform how we conceive of emergence in the specific case of human symbolic communication. Findings from this research will be analyzed and discussed in terms of interpretation and conventionalization, the two factors underpinning symbols identified in section 3 above.

Evidence related to the issue of interpretation can be found in experimental sign research as well as the natural development of new sign languages as observed in the field. Experimental sign studies show how signs are created, used and change over time under various conditions and parameters. These also reveal how modality of expression and communicative dynamics affect sign interpretations, and offer insights into the transition or lack thereof from nonarbitrary to symbolic signs. New sign language research in deaf communities documents aspects of the transition from an iconic and indexical gestural communication system to conventional sign languages in naturalistic settings. How these systems change over time, the differences in modality to vocal communication and how the resulting languages differ from typical spoken language will help to understand the forces that would have affected a communication system that shared many of these features in human history.

The issue of conventionalization is also addressed in experimental sign research and new sign languages. Experimental studies offer observations of transmission dynamics in real time on a small scale, while new sign languages show broader trends on a larger scale. In addition, a number of computational models have simulated the conventionalization of symbol systems. The success or failure of these models in constructing shared vocabularies reveals key factors that enable or impede the conventionalization process. Central to this process is the ability of recipients to identify the intended meaning of a novel arbitrary sign – that is, for the symbol to be *grounded*. Thus, an explanation of symbols must include the means by which novel symbols would have been grounded, and thereby conventionalize.

From this research I identify processes and factors involved in emerging sign systems generally, which are then applied to the context of human communication and evolution, the topic of the following chapter.

5.2 Chapter 3 - Human symbolic communication and emergence

In order to answer the question of how symbolic communication came about in human history, we must begin with an accurate and complete description of symbolic communication as it is carried out by embodied humans in lived experience. While in theory there are a vast number of different ways meaning can be shared through representations, humans engage in a specific version that is uniquely dictated by their bodies and perceptions. These specific features will determine what kinds of signs are created, how they are used and if and how their semiotic statuses undergo changes. Therefore, it is necessary to take into account the number and qualities of the modalities through which communication takes place, how those are utilized in communication and any potential interactions between the modalities. Incorporating this information into the investigation makes it possible to address the question of symbol origins in terms of embodied human experience and evolution.

In this chapter I review research on human communication, including the semiotic capacities of the vocal and gestural modalities, the way those modalities are utilized to accomplish symbolic communication today and the phylogenetic origins of this features. Symbol communication as practiced by humans is shown to be a multimodal, heterosemiotic system of unified symbolic vocal speech and nonarbitrary bodily gesture. Because symbols are not used in isolation and are coordinated with other sign types in this way, the question of symbols becomes how this particular mode of coupled symbolic-and-nonarbitrary communication emerged.

With the question fully specified and the relevant explanatory factors established, I then present and critique existing theories of symbol origins that begin from similar premises regarding bodily mimesis and nonarbitrary communication. These theories, in aggregate, are shown to not fully address the critical factors identified in this thesis and/or to rely on evolutionary processes inconsistent with the evolutionary principles

outlined above – namely, invoking the advantages of symbols as reason for their development. I argue that these issues can be resolved and explanatory gaps bridged by applying the information and ideas reviewed here – specifically, that the interpretation of vocalizations as symbolic and their subsequent conventionalization is a potential in-built consequence of the presumed communicative dynamics preceding symbolic language. Vocalizations would be interpreted as symbolic due to their semiotic profile relative to gestures, and conventionalization of arbitrary vocalizations would be enabled by the communicative support provided by accompanying nonarbitrary gestures. In this way, appeals to unviable evolutionary mechanisms are avoided and the modern state of multimodal, heterosemiotic communication is accounted for. Thus, the explanation for symbol emergence presented here is more satisfactory in terms of evolutionary principles and consistent with the available evidence. The following two chapters describe experiments to test hypothesis related to these claims.

5.3 Chapter 4 – Experiments: Interpretation

The arguments put forward in Chapter 3 make certain predictions regarding how vocal and gestural signs will be interpreted in relation to each other in the presumed context of symbol emergence. I present a series of experiments conducted to determine whether the modalities are processed in an integrated way like that observed in language studies today when the communicative and semiotic roles differ from that of today. Results indicate that arbitrary vocalizations are interpreted and learned like symbols even when they are communicatively uninformative and in the absence of the expectation to acquire a conventional symbolic system, providing support for the central claim of this thesis.

5.4 Chapter 5 – Simulations: Conventionalization

The arguments made here also make certain predictions regarding the ability for gestures to ground novel symbols. Building upon previous computational models that address the subject, I present a set of models designed to simulate multimodal heterosemiotic communication and test the capacity for information like that provided by nonarbitrary gestures to ground arbitrary signs. These simulations show that

accompanying nonarbitrary signals enable populations to converge on shared vocabularies across a range of conditions. These results support the claim that accompanying nonarbitrary gestures could ground arbitrary vocalizations and allow conventionalization to take place through natural communicative and cultural processes.

5.5 Chapter 6 – Conclusion

In the concluding chapter I will briefly review the research covered in this thesis, restate its central claims, describe its theoretical and empirical contributions to the field and suggest opportunities for related further research.

CHAPTER 2

Emerging sign systems

1. Introduction

The discussion of semiotics in the previous chapter showed that symbolic signs are forms that are interpreted as having an arbitrary relationship with the meaning to which they refer. This relationship is distinct from icons and indices, which are both nonarbitrarily connected to their meanings. To understand how symbolic interpretations of signs could arise from nonarbitrary signs, then, we should identify the causes, mechanisms and processes that instigate a transition from one interpretation type to another. It may be that certain social dynamics or transmission processes lead to the interpretation of signs as symbolic while others impede or prevent a transformation. Understanding the factors involved and the conditions that are associated with the development of symbols generally will provide valuable evidence for the present goal of explaining the emergence of symbolic communication in human history.

This chapter will examine research on emerging sign systems, or circumstances in which new signs are created and used in communication. This research sheds light on the nature of novel signs, how form and semiotic status change over time, and how a population's shared use of signs changes over time. The present review will focus on three main areas: lab experiments with human participants, sign language studies and computational models. Findings from these three areas will be discussed in terms of the explanatory factors of interpretation and conventionalization identified in the previous chapter as key to understanding the emergence of symbolic communication.

2. Experimental sign research

Studies on novel signs produced by humans have been carried out in order to understand how communication is established and how newly created signs change through processes involved in communication. Participants are given a medium through which to communicate and are in some way restricted from using linguistic or other standardized signs. The most common design employs a graphical interface,

wherein pairs or groups of participants draw images to communicate a list of concepts or about a task the partners are engaged in (Fay, Garrod & Roberts, 2008; Fay, Garrod, Roberts & Swoboda, 2010; Galantucci, 2005; Garrod, Fay, Lee & Oberlander, 2007; Garrod, Fay, Rogers, Walker & Swoboda, 2010; Healey, Swoboda, Umata & Katagiri, 2002; Healey, Swoboda, Umata & King, 2007; Theisen, Oberlander & Kirby, 2010). Another approach does not include a medium of any kind specifically devoted to sending and receiving signs, and participants must communicate solely through actions they can perform in the environment of the task (Scott-Phillips, Kirby & Ritchie, 2009). In this design, participants must recognize the opportunity to communicate, develop strategies to represent meaning with their actions, and then go on to establish conventions for doing so. The present discussion will focus on the first approach, as the results offer more insights into the development of symbolic signs from other sign types; however, the second approach represents an interesting comparison to these other findings and will be addressed after they are reviewed and discussed.

Some of these studies were designed specifically to investigate if and how symbolic signs develop from non-symbolic signs through communicative processes (Fay et al., 2008; Fay et al., 2010; Garrod et al., 2007; Garrod et al., 2010; Theisen et al., 2010). While others did not address this issue directly (Galantucci, 2005; Healey et al., 2002; Healey et al., 2007), their findings reveal important features of emerging communication systems. This section will review these experiments and their methodologies, evaluate their findings and relate them to the question of the emergence of symbolic communication.

2.1 Description and findings

Graphical communication studies involve a pair or group of separated participants communicating with each other through drawings. Participants must communicate a pre-given list of concepts, like ‘computer monitor’ and ‘cartoon’, but they are not allowed to use numbers, letters or other linguistic devices or symbolic conventions. Some studies do not restrict drawing abilities beyond these constraints, with participants free to draw for an extended period of time and employ as much iconicity as they desire and the medium affords (Fay et al., 2008; Fay et al., 2010; Garrod et al.,

2007; Garrod et al., 2010; Healey et al., 2002; Healey et al., 2007; Theisen et al., 2010). Another paradigm puts additional limitations on the medium of sign production that reduces participants' ability to portray iconic features in their drawings by using an apparatus that distorts and transforms marks by simulating a scrolling notepad (Galantucci, 2005).

Galantucci's (2005) experiment also differs from the others in that messages to communicate do not come from a pre-given list of concepts. Instead, participants play a computer game in which they try to achieve certain shared goals. Discrete locations in the simulated environment, or 'rooms', are placed on a grid, with each assigned its own unique visual representation in the form of familiar images – for example, a triangle or flower. There are variations on the exact terms of play and number of rooms, but the task generally involves a number of rounds in which an item is randomly placed in one of the rooms, and to win points players have to find the item and end up at its location together. To succeed at the task, participants must communicate via the drawing pad in order to coordinate their movements in the environment. Because the constrained medium prevents them from doing so by directly portraying images associated with rooms, they must develop some other strategy to communicate effectively.

A final dimension in which these studies differ from each other is in having participants engage in interactive communication – repeatedly exchange drawings about the same concepts, with participants alternating communicative roles – to track how sign form changes over time (Fay et al., 2008; Fay et al., 2010; Garrod et al., 2007; Garrod et al., 2010; Theisen et al., 2010). In these studies, pairs or small groups of participants communicate a single list of concepts over a number of rounds and switch between acting as sender and receiver. Consequently, each person draws and matches the same concepts multiple times and will share these experiences with their partner or partners. All interactive communication experiments use an unrestricted drawing apparatus. The signs from the initial rounds of these studies will first be described along with the signs created in the other studies under the unrestricted design (Healey et al., 2002; Healey et al., 2007) as well as those from Galantucci's (2005) restricted design. Following that, the findings regarding sign changes over time will be reviewed.

In experiments with an unrestricted drawing medium (Fay et al., 2008; Fay et al., 2010; Garrod et al., 2007; Garrod et al., 2010; Healey et al., 2002; Healey et al., 2007; Theisen et al., 2010), it was found that participants' drawings were iconic and indexical. In Healey et al.'s (2007) study, musical pieces were represented either through pictorial depictions of concepts evoked by the piece or metaphorical iconic depictions that resembled the musical structure (the crossmodal underpinnings of this kind of iconicity will be discussed in Chapter Three, section 2). In studies on the effects of interactive communication (Fay et al., 2008; Fay et al., 2010; Garrod et al., 2007; Garrod et al., 2010; Theisen et al., 2010), drawings from initial rounds were fairly elaborate and included many features corresponding to features of referents.

Although signs created by participants in Galantucci's (2005) medium-restricted design differ from those in the unrestricted version in being less directly pictorial, they similarly appear to be nonarbitrary. Under these conditions, three different systems of signs developed: one based on numbering the rooms according to their location on the grid, one based on depicting the rooms' assigned visual labels in some way, and one based on the spatial layout of the game map. The manner in which each of these systems makes use of nonarbitrary features will be addressed in the discussion section below.

In addition to finding communication was initiated with nonarbitrary signs, experiments on the effects of interactive communication also investigated how signs changed over time and use (Fay et al., 2008; Fay et al., 2010; Garrod et al., 2007; Garrod et al., 2010; Theisen et al., 2010). Under these conditions, drawings start out relatively complex and elaborate but over rounds of exchange lose many of their features and simplify (See Figure 2.1 for an example from Garrod et al., 2007). From such results, it is claimed that arbitrary, symbolic signs develop from nonarbitrary signs through interactive communication.

Two methods have been used to determine the sign status of interaction-derived drawings. The first measures graphical complexity – based on the amount and arrangement of lines or 'ink' contained in drawings – from early and later rounds. The second method tests how well naïve outsiders are able to guess from their

drawings the concepts intended to be represented by interacting partners, again comparing early versus late rounds. Final drawings were found to be less graphically complex than early drawings, a result paralleled by a decreased ability for outsiders to guess final drawings compared to early ones. The observed simplification in form and reduced transparency for outsiders are said by the authors to indicate signs have become arbitrary and symbolic. In addition, partners converge on shared signs to communicate the same concepts, from which it is claimed that conventional, symbolic signs can come about through interactive communication (Fay et al., 2010; Garrod et al., 2007).

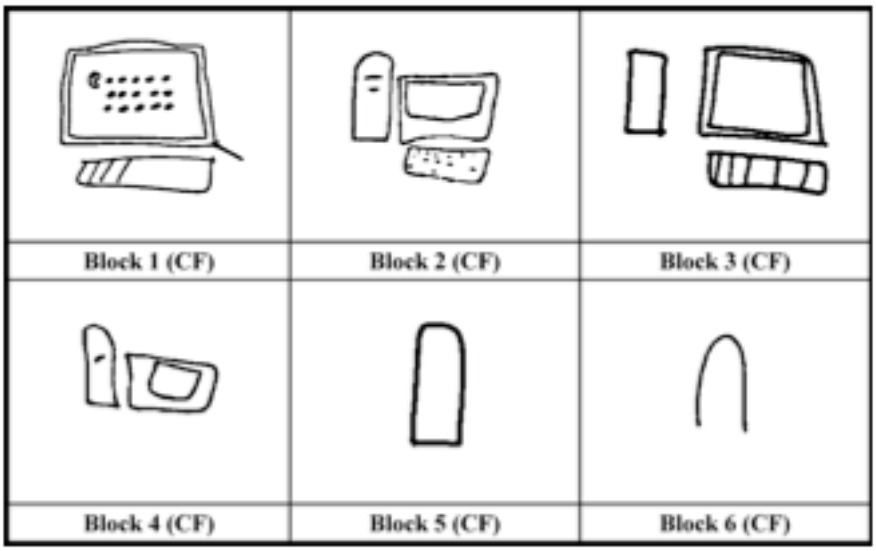


Figure 2.1. An example series of drawings for 'computer monitor' from Garrod et al. (2007; pg. 978). 'Block' number indicates the round in which a drawing was produced. 'CF' is a convention used in the paper for a particular experimental condition, the details of which are not discussed here.

2.2 Analysis

2.2.1 Nonarbitrary initiation

The first major finding of these experiments is that communication was initiated and established through nonarbitrary representations. It is first important to note that this conclusion is in fact an assumption, as none of the studies tested or documented how participants perceived and interpreted their signs. Nevertheless, there are reasons in this case to accept the conclusion as sound. For the unrestricted design (Fay et al.,

2008; Fay et al., 2010; Garrod et al., 2007; Garrod et al., 2010; Healey et al., 2002; Healey et al. 2007; Theisen et al., 2010), participants' successful communication without explicit feedback on their success or other means of understanding messages, together with the presence of many features of drawings that appear to correspond to features of the intended referents (see Figure 2.1 for an example), strongly suggests these features were recognized as iconic and/or indexical. The nonarbitrariness of signs is not as immediately apparent in Galantucci's (2005) restricted design, but a close examination of the three systems that were adopted reveals that these also likely relied on iconic and indexical representations.

First, the system based on depicting features of rooms' labels most clearly relies on nonarbitrary representation by portraying one or more distinct iconic features of those labels. For example, portraying a room's assigned shape by drawing the same number of marks that corresponds to the number of vertices in that shape. Figure 2.2a shows one pair's signs, which use three ticks to represent the triangle's three angles, five ticks to represent the flower's five petals, and so on.

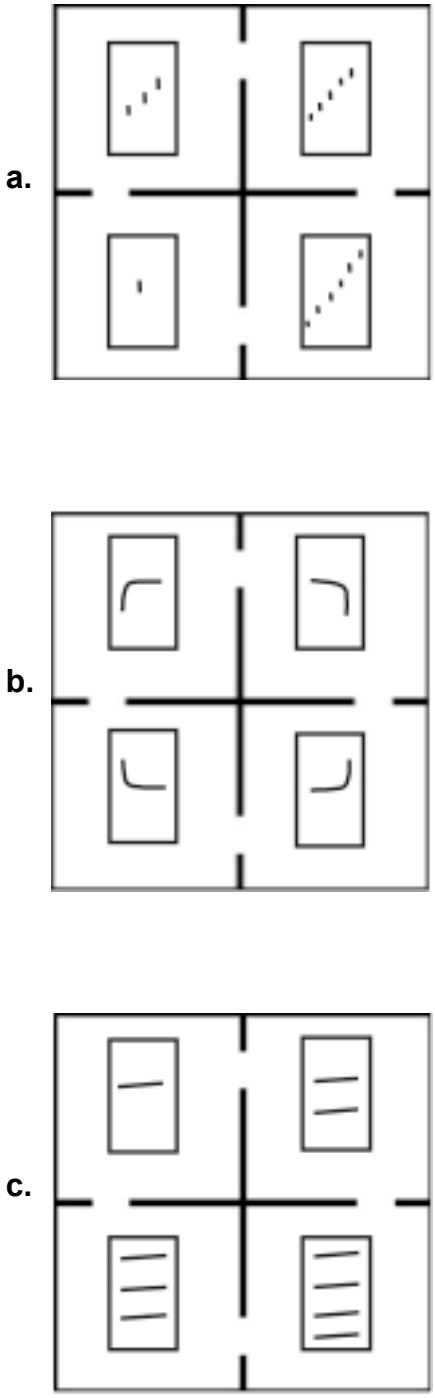


Figure 2.2. Example sign systems from Galantucci (2005; pg. 747). **a.** Drawings that represent the number of vertices in rooms' images. **b.** Drawings that represent the spatial layout of the game environment. **c.** Drawings that represent numbers assigned to each room.

Although mediated by prior knowledge of a symbolic system, or through reference to other structural aspects of the environment, the other two systems similarly rely on nonarbitrariness. First, the spatial system involved mapping the drawing space onto the layout of the game environment and using spatial relations between marks in the drawing space to refer to the different rooms. For instance, a right angle with its corner oriented to the top left of the drawing pad represents the room located at the top left of the grid (see Figure 2.2b). Thus, the organization of the drawing space is iconic of the organization of the game environment. Second, the numeration system involved assigning a number to each room, and making the corresponding number of ticks on the drawing pad to refer to a room by way of this number (see Figure 2.2c). This system is akin to Cartesian coordinates, though rooms were not necessarily ordered in the same direction. Numeral-coordinate signs are icon-index hybrids (see Chapter 1, section 3 for a discussion of sign-type blends), in that marks are iconic of the abstract quantities represented by numerals that were associated with particular rooms. This strategy is made possible by knowledge of symbolic numerals and the convention of numbering quadrants on a grid, but the way in which the information is communicated is via nonarbitrary signs.

Finally, a separate reason may explain why more minimal, less elaborately nonarbitrary forms were able to ground communication in this context. Unlike experiments with an unrestricted drawing medium, a consequence of the game play in Galantucci's (2005) task is that participants received feedback on their partners' intended meaning for the drawings they produced. In navigating the environment together and attempting to accomplish shared goals, participants could find out what a drawing was intended to represent through the course of play even if they did not correctly understand it when it was initially produced. In this way, feedback provided additional support to reduce uncertainty and establish the meaning of signs on top of the nonarbitrary information contained in the drawings' features.

The role of feedback as a support or substitute for nonarbitrary representation in establishing communication is further illustrated by Scott-Phillips et al.'s (2009) experiment mentioned above. In this design, pairs of participants played a computer game in which they have to move around on separate grids composed of four colored squares and end up on a square that is the same color as the one their partner is on. In

each round, both players are placed on new grids with a distinct pattern of colors (some of which are also on their partner's grid, some of which are not). Players can see each other's grids as well as each other's movements on the grids, but they cannot see the colors of their partner's squares. Players can move around for an unlimited amount of time, and once each thinks they are on the same color, they can choose to end the round. After a round has ended, players are told if they were successful, and the pattern of colors on their partner's grid is revealed. To succeed in this scenario, players must develop a strategy to represent specific colors as well as the presence or absence of colors. Given these conditions, the only medium available for participants to create signs is through their actions in the environment, which affords little, if any, opportunity for nonarbitrary representation. Nevertheless, some pairs successfully establish communication by designating distinct patterns of movement on the grid as representations of color information. This experiment is an example of establishment without nonarbitrary signs; however, feedback in the form of being able to see each other's color patterns after each round does the work of indicating for players their partner's intended meaning. It was also found that the perceptual salience of the color red led partners both to end up coincidentally on this color before any coordinated strategy was developed, which in itself helped to bootstrap communication. When red was removed from the color inventory in a follow-up experiment, the success rate of pairs dropped substantially. Shared perceptual bias is therefore also an important component for establishment in this case, but without the feedback at the end of a round, players would not be able to know that their partners also intended to end up on a red square. If this additional information were not available, the probability that participants could map meanings onto signals correctly is extremely low.

This blend of feedback and limited nonarbitrariness highlights an important aspect of emerging communication systems. For a novel sign to be understood and communication established, a receiver must be able to identify a producer's intended referent. In the absence of a shared symbolic system like language, as simulated in most of these experiments, people appear to rely on iconic and indexical representations to convey meaning. Information contained in the features of signs themselves allows for a receiver to interpret their meaning correctly without the need for much aid from additional information. The sign alone does not of course fully indicate meaning, as implicit information and inference from context are also

involved. Factors like shared situational awareness, intention reading and perceptual biases, which can be thought of as ambient information that colors and frames any instance of communication, are always at work when interpreting a specific sign in a specific context and will also aid in reaching the intended interpretation (the role of these factors will be more closely examined in section 4 below on computational models). These studies demonstrate that nonarbitrary signs are able to do much of the work in conveying meaning and reducing uncertainty, and thereby lessening the burden on inferential resources.

When nonarbitrary representation is not available, the work of connecting sign to meaning must be fully accomplished by some other route, such as feedback. Before discussing feedback further, it warrants mentioning that inferential strategies based on contextual factors in some cases can on their own do the work of grounding signs. This is part of some processes by which new symbols are spread to other users, a topic which will be addressed in more detail in section 2.2.3 below on conventionalization. However, in theory the same factors could allow the initiation of communication even if nonarbitrariness is not used or perceived. This is possible if the context of communication is tightly constrained, with a very limited set of potential referents, such as under experimental conditions. In more naturalistic settings it would require an information-rich communicative superstructure in which interlocutors are aware of each other's attentional states, goals and knowledge about the current situation. This level of contextual resources would support comprehension and ensure a correct interpretation is reached, even if no information is available from sign form or feedback. Although it may be possible for inference alone to enable the establishment of communication in this way, the necessary communicative environment presumably must result from preceding experience in which extensive shared knowledge and attentional focus has been acquired and shaped over time. It is difficult to imagine how such a situation could come about without previous communication of some sort, which brings the question back to the establishment of communication in the first place. Thus, it is likely that in the context of establishing communication, contextual resources alone may not be a reliable means to ensure novel signs could be understood without the aid of nonarbitrary representation.

When contextual information is insufficient, it is also possible for a sign system to develop without the aid of nonarbitrary information if communicators have access to feedback, which similarly serves to indicate the meaning of a sign. Feedback in this sense is when communicators are explicitly told, outside the communicative act itself, if a signal is correctly interpreted by the receiver, which indirectly indicates its meaning. Experiments that limit (Galantucci, 2005) or almost entirely eradicate (Scott-Phillips et al., 2009) the ability to utilize nonarbitrary signs critically depend on the presence of feedback for successful communication and the development of a sign system. In these experiments performance in the shared task and the information made available when players are told of their success or failure allows participants to infer if they correctly interpreted a signal and, usually, their partner's intended message.

Although feedback can accomplish grounding similar to nonarbitrary representation in this way, there are issues regarding its role in explaining symbol emergence for present purposes. The ecological validity of relying on feedback to bootstrap sign systems in natural settings is somewhat in question, as it rarely occurs in child language acquisition (Lust & Foley, 2004). Moreover, feedback essentially amounts to a disruption of the communication process; it is meta-information on the outcome of a communicative exchange that is presented apart from the exchange itself after it has taken place. Feedback can perhaps then be thought of as a proxy for nonarbitrary information, which is a direct and naturalistic mechanism to indicate meaning. Assuming communicators do not have direct knowledge of each other's intended meanings and must distinguish them from larger contexts of other possible meanings in the moment of communication, it seems that, in the absence of other mechanisms, nonarbitrary representations are the primary way to initiate successful referential communication. Including feedback in experiments on novel communication systems, then, is akin to an assumption that communicators would be able to indicate meaning or at least reduce uncertainty through nonarbitrary representations of some kind. The nature and role of nonarbitrary information in this process will be addressed further in section 4 below on computational models. For now, the point to emphasize is that when people must communicate without language, the default strategy is to do so by creating iconic and indexical signs.

One final point to note is that the success rate of participants in developing sign systems is lower in experiments that decrease or prohibit the utilization of nonarbitrary representation. Of the ten pairs in Galantucci's (2005) study, three pairs took over an hour to establish a system, two more pairs took over two hours to do so, and one pair never succeeded at all. The results of Scott Phillips et al. (2009) study are even more striking, wherein only seven out of twelve pairs successfully developed a sign system. In a follow-up experiment in the same study that removed the possibility for a strategy based on perceptual salience of colors, the number of successful pairs dropped to two out of twelve. In contrast, unrestricted nonarbitrary representation seems to be an effective and immediate way to establish communication and construct a sign system (Garrod et al., 2007; Fay et al., 2010). Thus, not only is feedback a less ecologically plausible mechanism, it is also not guaranteed to enable sign systems to emerge at all.

2.2.2 Nonarbitrary to arbitrary

Before discussing these methods and findings, it is important to note that based on the ideas presented above on sign interpretation, we must conclude that the status of these experimentally derived signs cannot be known with certainty. The perceptions of participants – whether they were relying on resemblance, concrete association or arbitrary mappings to interpret forms – were not examined or documented. Therefore, evidence of their status as made available in these studies is indirect and suggestive. Data obtained provide some indications of sign type, and from these experimenters have concluded drawings became symbolic. However, with an explicit consideration of sign interpretation in mind, alternative conclusions can also be reached from the same data. I will examine previous studies' analyses and the inferences made from their results and present an alternative analysis that reaches a contrasting conclusion. This conclusion will be based on the inferred interpretations of two separate, coherent perspectives in this experiment, those of interacting partners versus naïve outsiders.

Complexity measures will not be addressed in detail here, as they are wholly form-based and therefore largely uninformative for identifying sign type. Given the trends observed in the cultural evolution of writing systems, in which forms change and simplify over time (Olson, 1996), simplified sign form likely does correlate with arbitrariness, which is why this method may in some circumstances be a useful tool

for identifying symbols. However, they do not necessarily coincide, and it is the particular circumstances of these experiments that create precisely the conditions under which the two can become divorced. Why this is the case will become clear in the following discussion.

While checking the guesses of naïve outsiders appears to be a more perception-based method, it too proves to be problematic for effective sign assessment. Comparing outsiders' guesses to the meanings intended by partners with an interactive history amounts to comparing interpretations of the same form across separate contexts. Interacting partners have acquired unique common ground and awareness that an outsider does not share, and these differences constitute distinct contexts. It is in fact arguable that the drawings were not signs for outsiders at all, as they were not encountered as part of a communicative process. Communication is recipient- and sender-specific; messages are designed with a particular addressee in mind, and addressees interpret messages with regards to their producer's identity (Clark & Marshall, 1981). Drawings were not intended and tailored for outsiders, and outsiders were not aware of who produced the drawings and for what purpose. Because these essential aspects of communication are not present in the task of guessing, it is questionable whether we should consider outsiders' guesses as a coherent context of communication, which in itself would call into question the claim for symbols. Nevertheless, for the sake of argument, we will accept that the drawings were signs for outsiders. Even if we do so, however, there are reasons to conclude these were not symbolic.

First, consider the implications of blending sign interpretations from the two separate contexts of partners and outsiders. When assigned meanings for a form do not align across these contexts, it is taken by the authors to indicate the form is arbitrary. However, as the 'whiskers' example from the preceding chapter demonstrates (section 3.3.2), the relationship between form and meaning exists *within* a context. The meaning of the same form in one context has no bearing on its meaning, and therefore semiotic status, in another context. The reason the image from the example is symbolic of 'an underground international nuclear arms trader' is because the people who use the sign for this purpose have chosen intentionally to represent the meaning arbitrarily with this collection of lines. In contrast, the image functions as an icon for

'whiskers' in another hypothetical context because its users designed it to resemble whiskers, recognize this resemblance and use it to represent that concept. Therefore, in the same way, the relationship between form and meaning for signs developed through interactive communication exists in the interpretations of interacting partners. If we grant that the drawings were signs for outsiders, this relationship will similarly be determined by their interpretations. Although we do not have direct knowledge of participants' interpretations of signs, there is evidence to suggest forms were in fact perceived as nonarbitrary for *both* interacting partners and outsiders. The context-internal perceptions for each of these cases will be examined in turn.

Interacting partners who actively changed a form over time can recognize iconic features that remain in final stage drawings, and they are very likely relying on this recognition to interpret the signs. For instance, in the 'computer monitor' example from Garrod et al. (2007; Figure 1), it is clear from the sequence of drawings that the final image depicts a highly salient and easily recognizable feature of the second and all subsequent drawings: the computer tower. The drawing at this stage includes the depiction of an associated object and not the referent itself, but this simply means the sign has become indexical and iconic, by virtue of reference to earlier drawings of the same referent. The fact that partners maintain an unbroken chain of reproduced previous iconic features strongly suggests they are recognizing that iconicity, and therefore interpreting drawings as nonarbitrary signs.

On the other hand, for outsiders, who are asked to match drawings to a list of concepts, it is useful to consider what this task involves. Given no evidence to the contrary, we must assume that guesses are made by searching for iconic features in drawings that match features of referents in the list of concepts provided (in fact, it is unclear what other strategy would be used). Guessers therefore would perceive a resemblance (and possibly an association) between the forms and the meanings to which they choose to match them. For example, we can imagine that an outsider encountered the final 'computer monitor' drawing in Figure 2 and searched the list of concepts for what it might look like. One concept in Garrod et al.'s (2007) list is 'parliament'. The shape of the drawing is somewhat tower-like, and this outsider could see the image as resembling Big Ben, and therefore as representing parliament. If guesses are based on resemblance between form and chosen meaning in this way,

then outsiders are also interpreting signs as nonarbitrarily representing meanings. The fact that outsiders perceive *different* nonarbitrary features than the drawings' creators is not equivalent to arbitrariness, but rather independent icon- and/or index-based interpretations from separate contexts.

As laid out in Chapter 1, the features that distinguish and define symbols are i) that users interpret forms as arbitrarily representing meanings and ii) users share those arbitrary form-meaning mappings. However, in these studies, according to the alternative analysis presented here, neither of those cases holds (see next section for a discussion of point ii). Thus, it is not clear that the conditions simulated in these experiments are major forces in spurring the development of symbolic communication. This analysis also indicates that users do not appear to create symbols intentionally – either *de novo* or by transforming their own previously established nonarbitrary signs – which further
supports the position outlined in Chapter 1 that an explanation for the emergence of symbols should avoid appealing to insight and deliberate invention.

The low success of outsiders to guess interactors' intended referents may not reliably indicate sign type, but it does indicate something relevant to sign interpretation, what I will call transparency in form. Transparency in form – or outsiders' ability to guess how a form is used as a sign in reference to a particular context of communication – is an indicator of the likelihood that *intended* nonarbitrary features will be perceived (in Sonesson's (1997) terminology, a form that is transparent exhibits primary iconicity). In other words, guessability can be thought of as the likelihood for a form to be misinterpreted with respect to its users' intent. Low-guessability nonarbitrary signs are perhaps 'potentially symbolic' for non-users, though the question then becomes how outsiders are exposed to a form, perceive it as arbitrary, understand it is a sign nevertheless and then come to interpret it as representing the same meaning used by its creators (or perhaps some other meaning). In order to see how transparency in form could lead to symbols, it is likely interaction-derived signs must somehow be transmitted beyond the individuals that participated in their development, which would require additional social dynamics on top of interactive communication alone. Thus, transparency in form tells us something about the possibility for symbolic

interpretation to occur, but it does not demonstrate that symbolic interpretation *has* occurred.

2.2.3 Conventionalization

If the prediction that interaction-derived signs are not symbolic is correct, it not only raises issues with conclusions regarding how arbitrary signs are created, but also claims that conventionalization of symbols occurs as well. It may be that the conventionalization of still-nonarbitrary signs observed in these studies is not representative of the process of conventionalization for symbols. Because the mechanisms underlying the interpretation of nonarbitrary signs differ from those of symbols, the processes of transmission and reaching shared interpretation may similarly differ.

Nonarbitrary signs, given sufficient transparency of form and/or interactional history, are interpretable from their form and implicit sensory and associational knowledge. Correctly understanding the intended referent of a particular nonarbitrary sign created by an individual in a particular context will of course require some inference from additional details of the context of communication. Nevertheless, this process is largely accomplished by information contained in a sign's form – in other words, successful interpretation is somewhat guaranteed by the nature of the communicative act itself. Arriving at shared use of nonarbitrary signs, then, simply requires communicators to imitate and replicate signs they have understood by virtue of direct form-meaning relationships. This kind of imitation appears to occur with nonarbitrary signs in interaction experiments, and other studies have shown that speakers with a pre-established symbolic system who are engaged in a dialogue will similarly align in their use of linguistic structures (Garrod & Pickering, 2009). Thus, the imitative element underlying conventionalization of signs – whether nonarbitrary or symbolic – seems to result from a general feature of human social psychology. Given that, the aspect of the conventionalization of novel arbitrary signs that requires explanation is how these are first correctly interpreted. Section 4 below, covering computational models of symbol emergence, will explore this issue in more detail. Presently we can note that the reviewed experiments, under the current analysis, do not demonstrate with certainty the conventionalization of novel symbols, as they do not involve the use of arbitrary signs.

2.3 Summary and discussion

The central finding of experiments on emerging sign systems is that communication is initiated by way of nonarbitrary representations. If the ability to create such representations is not available or substantially restricted, some other mechanism is needed to ground forms in meanings by allowing communicators to access each other's intended messages; however, the ecological validity of this latter route to signs is questionable, and therefore less relevant to the emergence of symbols in human evolution. Following that, we can reason that, in naturalistic settings, the relationship between a form and meaning is first established through iconicity and indexicality based on shared sensorimotor knowledge (together with inference from ambient information) before other communicative forces could cause that relationship to shift to an arbitrary interpretation of signs. Thus, we can conclude that symbolic representations, when they do arise, somehow develop out of the nonarbitrary representations created to enable the initiation of communication. This is in line with the hypothesis of bodily mimesis, which holds that the transition from nonhuman communication to symbolic language would have involved an intermediate stage of nonsymbolic, referential communication, from which symbols emerged.

It is also important to note that, in the absence of additional supporting mechanisms, nonarbitrary signs shown capable of effectively establishing communication are representationally potent. That is, their forms typically embody many iconic features and robustly portray and/or indicate their referents. As the opportunity for nonarbitrariness is diminished, successful communication relies more and more on feedback. While these experiments preclude pointing, this strategy is similarly potent, as referents can be indicated with little uncertainty. The capacity for rich nonarbitrary representation in establishing communication is central to arguments presented in this thesis and will be revisited in Chapter 3.

In addition to lending support to the bodily mimesis hypothesis, these studies also tell us something about the potentiality of a transition from nonarbitrary to symbolic communication. As argued here, graphical communication experiments show the persistent, continued use of nonarbitrary signs under multiple communicative conditions, including different mediums of expression, purposes motivating

communication and social dynamics. If this conclusion is accepted, it suggests symbols are not guaranteed to arise in a given situation even if the potential exists, in this case by virtue of users that understand and use symbols regularly. Despite the fact that interpretations of signs as arbitrary can 'fall out' of communication through signs generally in modern communicative environments (Chapter 1, section 3.2.3), and despite the known advantages of symbols (Chapter 1, section 3.3), it appears that some circumstances impede or at least do not promote a transition from nonarbitrary to symbolic signs. In addition, it appears users do not deliberately invent symbols, further demonstrating the need to look for other forces apart from intention and insight. Thus, it is important to identify the factors that could in fact cause or allow a transition to take place.

Under certain conditions iconicity and indexicality are powerful and effective means of communication, and it seems people will continue to employ these strategies even though they have prior knowledge of symbol use and the possibility for symbol creation exists. For such nonarbitrary signs to become symbolic, their forms must somehow become disconnected from their meanings in the mind of some user or users. If a transition does take place, it will therefore likely occur unintentionally, resulting from a misinterpretation on the part of communicators. Because, at least in modern settings, a symbolic interpretation of an intended nonarbitrary sign involves a *mis*perception of iconic or indexical features, if and how a transition occurs will depend on the semiotic profile of the medium of expression and the social dynamics of communication. If a medium has a high capacity for iconicity and multiple features of a referent are depicted in a sign, the more likely it is that those features will be recognized and the intended nonarbitrariness of the sign preserved. If communicators have extensive common ground and shared understanding of the communicative circumstances, the more likely they are to correctly interpret each other's signs as intended. Thus, the interaction of these two factors in a given circumstance will influence whether symbolic interpretation arises.

The conditions of experiments on interactive communication do not appear to have spurred such a transition, which probably results from a combination of both signaling medium and transmission dynamics. The graphical medium is easily exploited for iconicity, and participants make use of this capacity to establish communication.

Although many iconic features produced in early rounds are lost in later ones, the shared history of interacting partners ensures recognition of those features that continue to be portrayed in drawings. These dynamics result in low transparency in form, as described above, though a break in the nonarbitrary connection between form and meaning for sign users likely has not yet taken place, and this transition may result from the additional involvement of outsiders in communication via interaction-derived signs. Even if outsiders would potentially interpret reduced signs as arbitrary, however, it is not clear from these studies how such signs move beyond the limited context of interaction. Imagine that after an interactive exchange like those described above takes place, one member of the pair or group is asked to communicate the same concepts to a new person. The person from the interactive pair knows that an outsider who did not witness their development will not understand final stage signs used with the previous partner, and would therefore not use them to communicate with the new individual. Instead, we expect a reversion to more elaborately iconic forms that can be understood without prior knowledge. Thus, it is not clear how interaction-derived signs would be used with outsiders, and therefore also not clear how transmission would occur. Precisely how interactive communication combined with other transmission processes could result in initially nonarbitrary signs becoming symbolic is likely an important and potentially fruitful line of research into the origins of symbols. Some preliminary findings from a very recent study (Caldwell, under review) suggest that a combination of interactive communication and vertical transmission to newcomers may lead to symbolic interpretations of signs, although this design also involved direct feedback, so it is unclear if and how this factor influenced the transition.

The argument that these studies also do not demonstrate conventionalization of arbitrary signs reveals how the different interpretive processes underlying nonarbitrary signs and symbols have implications for how signs come to be shared. Namely, the transmission – and by extension conventionalization – of novel symbols requires mechanisms for correctly interpreting signs that are not necessary for – or at least less critically involved in – nonarbitrary communication. The emergence of symbols from nonarbitrary signs, then, involves two kinds of grounding. First, communication is established and signs grounded by way of nonarbitrary relationships between form and meaning. A transition to symbols breaks this

relationship, and their forms must be grounded in meaning through other, additional mechanisms. I will refer to the initial grounding of communication as *establishment* and reserve *grounding* for the further establishment of shared arbitrary signs, as this is how the phenomenon is commonly referred to in other contexts. The question of grounding will be addressed in section 4, and the roles of both establishment and grounding in the emergence of symbolic communication will be discussed at the end of the chapter.

While the studies examined here may not demonstrate an actual symbolic transformation, it is of course still possible for initially nonarbitrary representations to become arbitrary and conventionalize through other or additional means. This process appears to have occurred in modern sign languages (Frishberg, 1975), and the next section will examine research on newly developed sign languages in order to understand how it could take place.

3. Developing signed languages

The study of established signed languages like American Sign Language as a mode of communication has shown that gesture can be used symbolically and function as a fully linguistic device, in contrast to the primarily nonarbitrary role in which it generally operates when accompanying speech (a topic covered in Chapter 3). The discovery of deaf communities in which people did not share an established sign language and new sign systems are in the process of development has shed light on how gestures could become symbolic. The findings of this research will be reviewed briefly, followed by a discussion of its implications for the question of symbol emergence.

3.1 Summary of findings

Research on deaf communities found that in situations where people do not share an established sign language, they nevertheless engage in communication through iconic and indexical gestures. Over time these initially nonarbitrary gestures take on all the properties of language, including arbitrariness and syntax, among other features (Meir, Sandler, Padden & Aronoff, 2010; Sandler, Meir, Padden & Aronoff, 2005).

This transition is observed by comparing the signs of different cohorts, or generations, in a system's development. In contrast to the signs produced by later cohort members, the gestures of signers from early stages of a new language appear to be overwhelmingly iconic and are more elaborately so, involving larger and longer movements of the body depicting a greater number of a referent's features. While much iconicity seems to remain in later stages, signers also use many gestures that appear to be symbolic.

Meir et al. (2010) describe how the emergence of new signed languages has been observed in two types of settings: i) small, tight-knit and fairly isolated groups, and ii) larger, more diverse communities with more exposure to outside influences. Small-group languages usually arise in isolated villages with an above average occurrence of congenital deafness, and therefore a much more concentrated level of deaf individuals in the population. Due to the high proportion of signers and the dense, multiplex social relations characterizing these communities, hearing individuals frequently will also learn and use the sign language. Large-group languages arise when many deaf individuals from different backgrounds are brought together, often in an institutional setting such as a school. Al-Sayyid Bedouin Sign Language (ABSL) exemplifies a small-group origin language (Senghas, 2005), while Nicaraguan Sign Language (NSL) emerged in a large-group setting (Senghas, 1995).

Meir et al. (2010) show that differences in group size and dynamics influence the development of sign systems and the resulting languages along a number of dimensions. Large-group derived languages typically have more extensive grammatical marking and less lexical variation than small-group languages. The more systematic structure of large-group languages is thought to be related to the social and communicative demands that arise when many people without common backgrounds must communicate with each other; when less understanding is shared through common cultural knowledge and experience, information must be more explicitly expressed in communication (this argument is akin to Wray and Grace's (2007) claim that language structure is related to the social structure of its speakers). There is also generally more contact with outside influences and other sign languages in large-group settings. Because individuals in small-group settings are typically more intimate and share a great deal of knowledge and experience, much of what is

communicated will not need to be explicitly encoded in the signing system. In addition, due to the relative isolation of these communities, there is very little influence from other cultures or sign languages. Meir et al. (2010) argue that the features of small-group settings are more naturalistic and representative of the context in which languages first arose.

Another major finding regarding small-group languages is the prevalence of lexical variation, or the use of multiple forms to represent the same meaning (Meir et al. 2010). For example, 70% of signs have three or more variants in one small-group derived language, Providence Island Sign Language (PSL), and many common lexical items have two or three variants in ABSL.

3.2 Analysis

New signed languages like ABSL and NSL illustrate that iconic gestures can undergo reinterpretation to become symbolic and shared by members of a community. These studies do not directly address how signs are deemed to be symbolic, though we can probably presume this in part involved asking signers about their own perceptions of signs. Nevertheless, there is strong evidence to suggest these claims are true and signers interpret many gestures as arbitrary.

First, purportedly symbolic signs show low what has been called here transparency in form (section 2.2.2); the meaning of gestures for non-users is not readily interpretable from features of sign forms. Even with knowledge of intended referents, a nonarbitrary relationship between form and meaning is not immediately obvious, which increases the likelihood that sign users, like non-users, do not perceive such a relationship. As argued previously, transparency in form is not a definitive indicator of a symbolic relationship, though in many cases the two are correlated. The reason it is an unreliable method for interactive communication experiments described above is because the experimental conditions create the precise circumstances under which nonarbitrariness would be intended and perceived in minimal, simplified forms – a situation in which signs are used by those who both created them and witnessed the loss (and retention) of some nonarbitrary features. In contrast, for developing sign languages many, if not most, sign users in later generations were not participants in a

sign's creation and most of its development, and therefore not privy to the historical knowledge that would allow remaining nonarbitrary features, if they exist, to be perceived. While still not definitive, low transparency in form in this particular case much more strongly suggests signs are symbolic for users.

Second, as mentioned above, many signs show high levels of lexical variation, or low levels of conventionalization. Importantly, signs that show this kind of variation are precisely those that are considered iconic. Because these gestures are understandable by recognition of nonarbitrary form-meaning relationships, they do not need to be conventionalized in order to communicate effectively. Signs that are considered symbolic do not show the same kind of variation, and this higher level of conventionalization likely results from an arbitrary form-meaning relationship, which demands shared cultural knowledge between users to function effectively.

Finally, the lexicon of the most recent cohorts contains words that are not found in earlier cohorts and refer to meanings that are not as amenable to nonarbitrary representation. For example, members of the most recent NSL generation have a sign for 'believe', a meaning for which the earliest cohort does not have a sign (Pyers & Senghas, 2009). The abstract, second-order nature of a concept like 'believe' could account for its absence in earlier stages, and its creation by later users was likely enabled by their having access to an established, at least partially symbolic, sign system. Furthermore, because it is not immediately apparent how best to portray the meaning of 'believe' iconically and/or indexically, it again decreases the likelihood that users perceive its sign as nonarbitrary. Although each of these reasons individually is not unequivocal evidence that purported symbols are in fact symbolic, taken together in combination strongly suggests we can safely accept the conclusion that some gestures in emerging sign systems transition from nonarbitrary to symbolic.

An important finding from this research is the fact that the transition from nonarbitrary, pantomime-like communication to symbolic, language-like communication spanned multiple generations.[6] The changes that resulted in arbitrary

[6] The term 'generation' is used here in the broad sense to denote distinct groups separated by time. For ABSL and other village sign languages groups correspond to the more specific sense of adult-child, family generations. NSL generations, on the

signs, grammatical marking and other linguistic qualities arose when new groups of signers were exposed to the founders' and previous generations' communication system. This is in line with the conclusion reached above regarding experimental sign research (section 2) that transmission to outsiders who were not a part of a sign's creation and development is likely a key causal element behind an established nonarbitrary sign becoming symbolic. Like the simulated communities in Fay et al.'s (2008) study, members of founding populations have direct or indirect knowledge of signs' development. When new individuals who were not involved in this process join the community, their acquisition of the system serves to sever signs from their historical origins, allowing new interpretations of form-meaning relationships to take place. As further evidence to the role of generational transfer, we can make a direct comparison and look to novel sign systems that do not get passed on to new users. Homesign is a term used for the gestures developed by deaf children born to hearing parents who do not use or learn a sign language (Goldin-Meadow, 2003). The signs used by children to communicate with hearing individuals are not adopted by others, and these gesture systems never fully take on the linguistic qualities found in developing sign languages (Goldin-Meadow, 2003). Thus, in naturalistic settings wherein a population creates nonarbitrary gestures to establish communication, it appears transmission to new individuals is a critical mechanism in transitioning from nonarbitrary to symbolic interpretations of signs.

While generational transfer broadly can be pointed to as a mechanism behind such a transition, the precise circumstances that lead intended nonarbitrary signs to be interpreted as symbolic remain unclear. It may be that child acquisition is a major factor, as new generations in NSL and ABSL are made up of either children entering the school or deaf children born into the village. Because children lack much of the knowledge and experience held by adults, they will not recognize as easily all the nonarbitrary features that an adult might perceive. However, transmission to children is not a theoretical necessity, at least in the modern communicative environment, so it is also possible that certain social and communicative dynamics instigate symbolic interpretations of signs. Although child acquisition may not significantly affect or change the process of symbolic reinterpretation itself, it could speed the process of

---

other hand, are new classes of students entering the school and learning signs from more senior students.

conventionalization of new symbols by increasing the number of symbolic interpretations held in the population that could potentially spread to others.

Similar to the question of reinterpretation, the precise details underlying the conventionalization of newly interpreted symbols are not apparent from this research. We can reason that some other communicative supports apart from transparent form-meaning relationships allow arbitrary signs to be understood and adopted by new users. Because developing sign languages arise in natural settings, direct feedback likely plays a minimal role. The following section on computational models explores other possible mechanisms, and the details of both reinterpretation and conventionalization will be addressed in the general discussion at the end of the chapter.

Although developing sign languages offer important insights into the emergence of symbolic communication, we can note some interesting caveats that suggest this phenomenon might not exactly mirror how the process took place in human evolution. First, these languages arise in a rich setting of symbolic culture and communication. Symbolic representation is ubiquitous in the larger cultural milieu of modern societies, and its prevalence may in some way support and structure developing sign systems. Deaf individuals are fully aware that hearing people communicate with each other through sound, images and orthography, much of which would presumably appear arbitrary, and therefore symbolic, to an observer. This knowledge of symbolic communication and culture likely influences the expectations and cognitive mindsets of signers, which in turn could effect how signs are interpreted, learned and transmitted. While these factors may not be solely responsible for spurring symbolic interpretations of signs, they may affect a transition to conventionalized, symbolic communication in other ways. One possibility is that the process takes place more rapidly, though there may be more subtle affects that are difficult to predict.

A second distinction worth noting is that sign languages – both new and institutionalized – retain a great deal of iconicity and/or indexicality (Perniss, Thompson & Vigliocco 2010). Many lexical items appear to be based on nonarbitrary features, and even some grammatical relations are expressed through iconic use of space (Meir, 2002). Spoken language also makes use of nonarbitrariness (a topic that

will be examined more closely in Chapter 3), though it seems to be to a lesser extent than signed languages. This persistent and pervasive use of nonarbitrary representation is related to the relatively high degree of lexical variation mentioned previously. Conventionalization is especially low in new signed languages, but even well-established, highly institutionalized signed languages show considerable lexical variation. For example, 28.6% of signs in American Sign Language have three or more variants (Washabaugh, 1986; cited in Meir et al., 2010). To understand this it is important to note that lexical variation as used by the authors appears to be neither the same as sociolinguistic variation like phonological or phonetic differences across dialects nor equivalent to synonymy. For example, Meir et al. (2010) describe how the word *cat* in ABSL has three versions, and each depicts a different set of cats' features – one involving whiskers, another footprints and the third the action of a cat licking its front paws. The forms for these three signs bear little or no resemblance to each other, but nevertheless all resemble the same referent in some way. Thus, this kind of variation is not analogous to accent variation, wherein related social groups pronounce what is recognizably the same word with distinct yet similar sounds. There are also reasons to conclude the phenomenon is not simply synonymy. Because these signs were elicited in an experimental setting, we can assume signers were given explicit, specific and basic meanings and asked for that word in their language. A context such as this leaves little room for nuances in meaning that could give rise to multiple potential translations for signers to offer. Instead, sign variation probably results from the fact that some referents have many salient features, and nonarbitrary representation allows for the same meaning to be conveyed by portraying multiple, distinct sets of features. From the evidence available, then, it appears these findings on lexical variation are genuine dissimilarities in conventionalization with spoken language, which suggests the process of conventionalization in the emergence of vocal symbols may differ from that of sign languages in important ways.

Reduced conventionalization and continued reliance on nonarbitrariness ultimately follow from perhaps the most fundamental difference in circumstances – modality of communication. As discussed in Chapter 1, section 3, the medium in which signs are conveyed will affect interpretive processes as a function of its capacity for different types of representation. The gestural modality allows for many opportunities to represent nonarbitrary features, and new signed languages demonstrate that people not

only exploit this possibility in order to establish communication but also continue to use nonarbitrary signs even after symbolic communication – a superior strategy in many ways (Chapter 1, section 3.3) – arises. This suggests that when the medium in which signs are conveyed has a high capacity for nonarbitrariness, symbolic reinterpretation may be somewhat limited. In other words, if nonarbitrary features are easily representable and recognizable, people will continue to use and correctly perceive nonarbitrary signs even if they have knowledge of symbols otherwise. In addition, the fact that the purportedly more advantageous strategy of symbolic communication is not more exploited by signers suggests the theoretical advantages of symbols are not especially relevant for communication in practice. This is further evidence that pointing to advantages of symbols may not be a satisfactory explanation for their emergence (Chapter 1, section 2.4).

In addition to taking place in a *different* modality, communication in signed languages is distinct in that it occurs in a *single* modality. Human communication is typically carried out in multiple modalities simultaneously – both vocal and gestural (multimodality will be discussed in detail in Chapter 1). The fact that signed languages are limited to gesture alone likely affects how both reinterpretation and conventionalization occur. The presence of two channels of expression means signs produced together in separate modalities will be perceived relative to one another, which will in turn determine the modality in which symbols arise. The fact that gestures become symbolic in developing signed languages may be largely a consequence of gesture being the only available channel for this transition occur, and therefore not necessarily representative of the process under normal circumstances of human communication. The following chapters will discuss these issues further, including the specific features of each modality, their interaction and how multimodality relates to the emergence of symbolic communication in human evolution.

3.3 Summary and discussion

Despite the fact that a number of details remain in question, research on developing sign languages reveals important large-scale features of emerging symbolic communication. In line with experimental studies, it shows that communication is

established with nonarbitrary representation. When humans who do not possess knowledge of a language need and desire to communicate with each other in natural settings, they do so initially by creating iconic and/or indexical signs. In this case, it seems the medium of gesture allows for a sufficiently rich and informative nonarbitrary communication system to develop that can further support the emergence of a different kind of representational strategy. Gesture is able to *establish* communication, from which symbolic interpretations arise through certain transmission processes that occur when naïve outsiders join the community and learn the sign system. In addition, communication through the gestural modality allows for these novel symbols to be *grounded* for others and conventionalize. The capacity for gesture to enable both of these processes presumably lies in its high potential for nonarbitrary representation, and this quality together with the unimodal nature of communication likely accounts for many of the differences between signed languages – both developing and institutionalized – and spoken language. The aim of following chapters will be to investigate the particular features of typical human communication and how the two processes of establishment and conventionalization would be realized in that context.

This and the previous section have been informative predominantly on the subject of the interpretation component of symbolic communication. For experimental studies, we cannot confidently accept results on conventionalization, as the presence of symbols remains in question. In addition, the conventionalization of novel signs is a group-level phenomenon and involves complex interactions between multiple individuals over time, and is therefore difficult to directly observe in natural settings. The following section will review computational research focusing on conventionalization that models this complexity and scale.

4. Computational models

We can first acknowledge which aspects of the emergence of symbolic communication computational simulations can help us to understand. Hutchins and Johnson (2009) argue that models implemented thus far assume prior knowledge of

intentional representation and therefore do not address the emergence of a symbolic medium or symbolic capacity, only the emergence of structure in a symbolic medium. In other words, it is left unresolved how the ability and/or inclination to use signs to communicate (be it through iconic gestures, arbitrary sounds or some other form) may have itself arisen. In the terms of this thesis, the criticism is that simulations of the kind conducted so far take for granted the interpretation element of signs generally – that communicators create and perceive intentional, referential representations – and only address the conventionalization of arbitrary signs in a population. This claim may be well founded, as models involve the direct transfer of messages between agents. However, we can accept models' assumption of interpretative abilities, as the state in the emergence of symbolic communication under examination here similarly assumes the ability for this type of representation through bodily mimesis, though what *kind* of interpretation users have of signs is of course relevant (how the capacity for bodily mimesis relates to a prior ability to interpret symbols will be addressed in chapter 3). While this means the computational models reviewed here cannot tell us how signs become interpreted as symbolic, they can nevertheless inform our understanding of how new symbols, once they come about, could spread throughout a population and conventionalize. Conventionalization in the context of simulations is when all agents in a population hold the same signal-meaning mappings, which is referred to as *convergence*.

Because the focus of this section is the conventionalization of symbols, it will be examining the concept of grounding discussed above (section 2.3). When a new arbitrary sign is used, its receiver must somehow correctly identify the meaning that sign was intended to convey. Successful grounding allows for particular arbitrary sign-meaning mappings to be understood and adopted by other individuals, and thus potentially spread throughout a population, leading to convergence. The following sections will explore the mechanisms underlying these related phenomena by first reviewing how the symbol grounding problem has been addressed by various approaches and then discussing different models of convergence based on these approaches.

4.1 Symbol grounding

The problem of symbol grounding follows from the fact that symbols, being arbitrary, have no direct or immediately knowable connection to their referents. As famously illustrated by Quine (1960), upon hearing an unknown word, there is no way to know with full certainty and precision its intended referent. That is, it is not immediately clear how communicators come to adopt shared word-to-meaning mappings because a given context of speaking in theory affords an infinite number of possible interpretations.

One area in which symbol grounding has been the focus is child language acquisition, which involves learning an established symbolic system of word-meaning mappings. Traditionally, abilities related to intention reading and sharing attentional states are proposed as mechanisms that allow children to identify correctly the target object referred to by utterances directed at them (reviewed in Smith, Smith, Blythe & Vogt, 2006). Computational and mathematical models have demonstrated that cross-situational learning strategies also enable the acquisition of vocabularies (Smith et al., 2006; Smith, Smith & Blythe, 2011). Cross-situational learning is when signals are mapped to meanings by eliminating potential signal-meaning pairs based on their co-occurrences across different contexts of communication. For example, say a signal, s_1, occurs in three different contexts consisting of three objects each, $\{o_1, o_3, o_6\}$, $\{o_1, o_2, o_3\}$ and $\{o_2, o_3, o_5\}$. By comparing the common contexts of s_1, one can infer that it refers to o_3, as this is the only object present across all contexts. Cross-situational learning is therefore one mechanism that can reduce uncertainty in vocabulary learning and helps to resolve the problem of symbol grounding in the acquisition process. Although cross-situational learning aids in mapping signals to meanings in this way, we can note that it is not strictly a grounding mechanism in itself and is more a further learning process that operates on the output of the grounding process. That is, it does not connect signals to meanings in the communicative exchange itself by indicating meaning but instead rules out competing pairings over the course of multiple communicative exchanges. Nevertheless, for present purposes we can treat cross-situational learning as a grounding mechanism as it accomplishes similar ends.

Research in artificial intelligence has also shed light on the phenomenon of symbol grounding (reviewed in Steels, 2003 and Roy, 2005). These studies demonstrate that communicating robots can coordinate mappings between signals and meanings that correspond to sensory information. Typically, members of a population of robots endowed with one or more sensory channels attempt to communicate with each other about features of their shared environment. Successful grounding relies on the ability of communicators to access each other's attentional states via perceptual information. In this way, signals become associated with aspects of perceptual content. The 'pragmatic feedback' (Steels, 2003) embodied in robots' interactions allows a hearer to correctly identify a speaker's intended referent by coordinating its sensory channels to focus on the same environmental feature to which the speaker is attending. The mechanism for establishing shared attention states has been implemented as direct access to a speaker's current sensory inputs (equated to pointing) (Steels & Kaplan, 1998) and eye-gaze tracking (Roy, 2005). These strategies partially mirror some aspects of the human strategies for establishing joint attention, though a number of other, less well understood mechanisms are also involved (Steels, 2003). Error feedback is another method of establishing reference, but it has been shown to be functionally equivalent to reliably precise joint attention (Vogt & Coumans, 2003), as both mechanisms serve to reduce the context of communication to one object.

The reviewed research indicates that symbols can be successfully grounded in meanings, either through situated, embodied interaction or a certain type of learning strategy. Successful grounding is one element that ultimately leads to successful convergence. Grounding can take multiple forms – be it cross-situational learning strategies or various methods of establishing joint attention. Different types of grounding will have differential effects on the extent of convergence in emerging communication systems. The following section will discuss the ways in which successful convergence has been attained in various models.

4.2 Models of convergence

A vast number of studies have been carried out that model the emergence of shared vocabularies. However, these can typically be classed into a few distinct categories based on the parameters of their designs. For example, a number of studies simulate

joint attention and/or corrective feedback strategies for grounding symbols (Baronchelli, Felici, Caglioti, Loreto & Steels 2005; De Jong, 2000; Steels & Kaplan, 2002; Steels & Vogt, 1997; Vogt 2001). Other studies test emergence using cross-situational learning strategies, though fewer of these exist (A.D.M Smith, 2001; Vogt, 2000a, Vogt, 200b). (A review of the literature can be found in Cangelosi & Parisi, 2002.) These strategies will be discussed in more detail below, but due to this overlap in factors underlying conventionalization, a comprehensive review of the modeling literature will not be carried out here. Instead, a selection of representative designs will be discussed and analyzed to highlight the critical issues of relevance to the present inquiry.

4.2.1 Parameters

Models of emerging symbolic systems vary along four main parameters: grounding mechanism, population structure/transmission dynamics, learning method and signaling method. Some of these parameters may overlap and interact. For example, learning method and signaling method (see below) may be in some ways related to grounding strategy, demonstrated most clearly in A.D.M Smith's (2001) cross-situational learning model. Nevertheless, it is possible and useful to separate these mechanisms to some degree in order to understand how shared symbolic systems emerge.

A grounding mechanism may be one of those discussed above. Precise joint attention serves to limit communication to a single signal-meaning pair, which will be referred to as 'perfect context reduction'. All models discussed use perfect context reduction except A.D.M. Smith (2001), which relies on cross-situational learning.

Population structure and transmission dynamics will dictate the total number of agents communicating with each other, which agents communicate with which other agents and if the members of a population change over time. The studies under review use either a *closed-group* or *replacement* method (Mesoudi & Whiten, 2008). Closed-group structures consist of a static population wherein all members communicate with each other with equal probability (i.e., horizontal transmission alone). The replacement method differs in that at regular intervals certain members of a population are removed and replaced by new members with no previous experience

communicating or learning a vocabulary. The replacement structure implemented by K. Smith (2002) simulates a specific type of iterated learning (Kirby, 1999), which is the process by which the output of one agent is used as the input of another agent, whose output is then used as the input for the next agent, and so on. While closed-groups involve iterated learning, in that as part of the process of building individual vocabularies agents produce signals that other agents receive as input, and those agents go on to produce signals for other agents, there is not a uni-directional, linear transmission structure. Transmission is entirely vertical in Smith's (2002) replacement groups; new members learn from signal-meaning pairs produced by existing members, but new members do not communicate back to existing members or with each other. In addition, new members in replacement groups will influence a developing signaling system differently than members that remain in a closed-group over time. Members of each new generation begin at an initial state of no communicative experience and learn from existing 'speakers'. Repeated learning by new individuals and the removal of agents with acquired vocabularies shapes a developing communication system by amplifying the effects of the particular learning biases agents possess (Kirby, Dowman & Griffiths, 2007). In contrast, it is the accumulated experience of a static set of agents in closed-groups that influences how their vocabularies are shaped over time.

Learning strategy determines how agents associate signals and meanings. The most basic learning strategy is simple positive association, wherein relationships between signals and meanings (whether measured in frequency counts or connection weights) are strengthened whenever they co-occur. The learning method modeled in most of these studies is a somewhat more robust one that includes negative disassociation, referred to as lateral inhibition. In lateral inhibition, connections between co-occurring signals and meanings are increased, while at the same time connections between those and other meanings and signals, respectively, are reduced. This type of learning functions as a bias for one-to-one mappings between signals and meanings – or a bias against homonymy. A.D.M. Smith (2001) models cross-situational learning, which involves tracking an inventory of potential signal-meaning pairs based on all previous contexts of individual signals experienced with multiple meanings. Agents come to map signals to individual meanings over time based on their co-occurrence and the co-occurrence of that signal with other meanings. This strategy involves

continually tracking and updating accumulated signal and meaning co-occurrences by eliminating previously stored potential signal-meaning pairs when contradictory evidence is found in new contexts. It is important to note that cross-situational learning tracks frequencies for *all* meanings in an observed context. Keeping track of every potential meaning allows a learner to use past experience to hone in on a correct meaning pairing by comparing the meanings experienced in each new context with all those in past contexts.

Signaling method in most models is generally straightforward and simply involves a speaker choosing signals based on strength of association with meanings. A.D.M. Smith's (2001) model again differs in this respect, as he employs a mechanism called an *obverter* (Oliphant & Batali, 1997) for signal selection. With an obverter strategy, agents design messages so that they themselves would understand them; for a given meaning, a signal is selected with the strongest co-occurrence relationship with that meaning based on their own communicative history. K. Smith (2003) argues that an obverter also amounts to a bias for one-to-one mapping.

In previous studies, successful convergence on a shared vocabulary depends crucially on the implementation of particular features specified above and/or a combination thereof. This review will look at three major paradigms and their prototypical models in order to show some of the different ways in which convergence can be attained.

4.2.3 Routes to convergence

The most basic model has been carried out with both robots (Steels, L., F. Kaplan, A. McIntyre, & J. Van Looveren, 2002) and simulated agents (Baronchelli et al., 2005; Vogt & Coumans, 2003; Steels & Kaplan, 1998). In this paradigm, simulations begin with a population of agents with the ability to produce and receive signals, but they do not start out with a vocabulary or share any sort of signaling system. Thus, agents do not initially have a way to correctly interpret signals by way of a common code, and instead must, through some other means, obtain knowledge of the signaler's intended meaning. All of these models provide grounding for signals in this way through perfect context reduction. In addition, all implement closed-group population structure and lateral inhibition learning. Given these conditions, after a period of repeated communication events populations attain high communicative accuracy (at

or above 95%), which indicates convergence on a shared signaling system. Vogt & Coumans (2003) showed that these results hold for population sizes of up to twenty agents. Steels & Kaplan (1998) also tested how removing the mechanism for joint attention affected outcomes by making the context of communication include the entire meaning space – in other words, testing perfect context reduction versus zero context reduction. Simulations in which context reduction was hindered did not result in successful convergence, demonstrating the mechanism is crucial for convergence in this paradigm (Steels & Kaplan, 1998).

K. Smith's (2002) model, which also implemented perfect context reduction, compared the effects of different learning strategies and iterated learning on the construction of shared vocabularies. As mentioned previously, the population structure was based on the replacement method, which removes some agents and introduces new ones into a population each generation. Instead of repeated bi-directional communication between the same set of agents, his model simulates repeated learning by new agents from 'utterances' produced by a subset of the existing members of the population. Agents that comprise a population at the start of each simulation begin with random associations between signals and meanings. Smith found that populations with no initial shared vocabulary can construct a shared system and reach convergence only when learning institutes lateral inhibition, which translates to a bias against homonymy. Subsequent work (Smith, personal communication) further showed that when the replacement method was not implemented the same results hold, indicating this particular population structure is not critical to convergence. Perfect context reduction is likely a critical element for successful convergence in this model, as it is what allows agents to associate the correct meaning with a given signal, though this was not specifically tested. However, K. Smith (2002) demonstrates that this grounding mechanism on its own is not sufficient. Under such conditions, a certain type of learning is required to shape vocabularies as they are acquired by agents in each new generation.

A.D.M. Smith's (2001) model is similar to those already discussed in that agents begin simulation runs with no shared vocabularies. Beyond that, it is distinct along a number of dimensions. First, communication events are not limited to a single signal paired with a single meaning; instead, 'hearers' receive a signal together with a

context of meanings (this parameter was varied, but the standard size was five). Agents use cross-situational learning to build their individual vocabularies over time through experience of multiple, overlapping signals and contexts. The obverter signaling mechanism also exploits structural knowledge of agents' lexicons and previous experiences with other agents to determine what signal-meaning pairs are to be produced by a 'speaker' in a given communication event. Smith found that populations consisting of two agents would eventually converge on a shared signaling system. Vogt & Coumans (2003) replicated Smith's model and tested the effects of varying population size. They found that – unlike models with perfect context reduction – populations of ten or more agents no longer reach high communicative accuracy, and thus do not converge on a shared vocabulary. Cross-situational learning and obverter signaling can therefore lead to the emergence of shared vocabularies, though under more limited conditions.

The reviewed models have identified some mechanisms that, given certain conditions, can reliably lead a population of agents to associate the same signals with the same meanings and successfully communicate with each other. In two of the three model types, some sort of joint attention-related device is critical for establishing shared reference (Steels et al., 2002; Steels & Kaplan, 1998; Smith, 2002). The third illustrates that a special kind of learning strategy together with obverter signaling can also lead to convergence, though its ability to do so begins to break down as the number of agents communicating with each other increases.

4.3 Analysis

These models demonstrate that, under certain conditions, arbitrary sign-meaning mappings can become conventionalized across a population. Successful convergence in these cases depends on a small number of critical factors, which can be described in terms of the parameters discussed above (section 4.2.1).

In two of the three paradigms, perfect context reduction as a grounding mechanism and lateral inhibition as a learning strategy appear to be key to reaching convergence. K. Smith (2002) showed that a bias against homonymy underpins convergence, even with perfect context reduction. Steels & Kaplan (1998) confirmed that absence of

direct access to intended meanings prevents convergence, even if agents possess such a bias. This learning strategy is relatively simple, and humans regularly exhibit it in linguistic settings (K. Smith, 2002); therefore, we can safely take this type of learning as a given for present purposes. The parameter of most interest here, then, is perfect context reduction, which we can attribute to some mechanism for establishing joint attention. Steels & Kaplan (1998) did not vary reduction by degree, so it remains unclear if convergence would be possible at smaller context sizes. Given the results, however, we can conclude that for these conditions, joint attention is critical to successful symbol grounding.

A. Smith's (2001) model shows that other mechanisms can ground symbols without the aid of perfect context reduction. However, its ability to support successful convergence is less robust than models that implement perfect context reduction, as demonstrated by Vogt & Couman's (2003) extended analysis. In addition, as mentioned above, cross-situational learning is not strictly a mechanism for grounding signs in the moment of a communication event, and is instead more a general learning strategy that supports the convergence process over time. Furthermore, Smith et al. (2011) found that human learners apply a weaker form of cross-situational learning when referential uncertainty is high and exposures to a word do not occur consecutively – conditions that presumably more closely reflect those of emergence. Given the less robust convergence results with cross-situational learning, its less direct role in individual communicative acts and its weaker application in real-world settings, joint attention appears to be the more powerful and relevant mechanism concerning the conventionalization of symbols. As such, the role of joint attention will be the focus of the remainder of this thesis.

If grounding by way of joint attention is the central factor underpinning conventionalization, it is by extension necessary to consider how it is achieved in human communication. A diverse suite of abilities are thought to be involved in establishing and maintaining joint attention, including gesture recognition, eye-gaze tracking, shared task awareness, script execution and recognition, and emotion recognition and synthesis (Steels, 2003). Therefore, in understanding how novel symbols could have been conventionalized in the context of human evolution, we should consider how these mechanisms operate in human communication.

5. General discussion

Research on emerging sign systems offers important information regarding the establishment of communication, the factors underlying shifts in sign interpretation and the processes by which novel arbitrary signs are transmitted beyond their original creators and become shared across a population.

In the absence of a common symbolic system, nonarbitrary representations allow people to express meaning and understand each other's messages. Direct or indirect feedback can also serve to bootstrap communication, though it i) is less effective (Scott-Phillips et al., 2009) and ii) has questionable ecological validity (Lust & Foley, 2004). Moreover, it appears the establishment of communication is not guaranteed when the medium of expression does not afford sufficiently large representational capacity (Galantucci, 2005). Therefore, signs that embody rich and robust nonarbitrary representation are likely needed to construct a communicative foundation, and it is from this foundation that new interpretations of signs can arise.

Once communication through nonarbitrary signs is successfully established in this way, the possibility for symbols presents itself, at least in the context of modern human communicators observed in experiments and the field. Intended nonarbitrary features in signs may not be recognized by others, who then interpret the sign as arbitrary, and therefore symbolic. As evidenced in experimental sign research and developing sign languages, it is a *combination* of form-meaning relationship and transmission dynamics that underlies the transformation from nonarbitrary to arbitrary. Complexity in form and extensive common ground tend to ensure nonarbitrary features are perceived, while simplicity in form and lack of shared knowledge increase the likelihood of misperception. The interaction of these two factors will determine if signs are misperceived and become detached from their nonarbitrary origins for an individual, opening up the possibility for shared symbols to emerge. In this way, how signs are perceived is ultimately responsible for determining if a sign becomes arbitrary, which indicates that we should look to and consider the effects of the particular features of the modality through which signs are communicated. That is, the physical and semiotic qualities of a modality and/or

modalities are causally relevant to the question at hand and can help to formulate our explanation.

Of course, misperception does not automatically lead to or guarantee this outcome. Signs must not only be misperceived, they must also be reinterpreted as symbolic. Furthermore, for symbolic communication to emerge in a population, individual reinterpretations must be used with and adopted by others. A consideration of the details of these two prerequisites will help to frame our explanation of how symbols came about in human history.

How might *mis*interpretation of a nonarbitrary sign result in *re*interpretation of the sign as symbolic? Assuming that communication is preserved and meaning is successfully conveyed, this process requires both skilled intention reading and some kind of contextual support. First, the signer's communicative intent itself must be recognized so that an interlocutor will attempt to interpret the sign in the first place. We might expect this prerequisite to hold generally, as such intention reading is the foundation on which the act of referential representation rests. However, this will also be affected by the number, qualities and relative communicative roles of the modalities through which expression takes place, when more than one are used simultaneously as in human communication. These issues will be addressed in the following chapter, but for present purposes we can reason that with an awareness of communicative intent in place, and after failing to perceive any meaningful features in the sign's form, a receiver might infer an intended meaning based on other context-based factors. If the receiver concludes the sign represents a meaning that appears arbitrarily related to its form, that sign has become a symbol for this individual.

The critical point in this process lies in correctly identifying (at least part of) signers' intended meaning despite misinterpreting their representations. In other words, because the nonarbitrary sign is arbitrary for its recipient, it must be grounded in information apart from form. Broadly, this grounding could be accomplished via two routes. One, proposed by Tomasello (2008), involves an observer witnessing others' successful nonarbitrary communication without recognizing their signs as such. The observer may not understand a sign as intended but could nevertheless infer its meaning based on the interlocutors' subsequent actions. In this scenario, form is

successfully mapped to meaning by virtue of *others'* correct interpretations. The other possibility is that a direct recipient could infer intended meaning based on other sources of information in the context. This possibility rests upon the ability to identify a signer's focus of attention, and therefore involves the same grounding mechanisms discussed above (section 4.1).

After a newly reinterpreted symbol has been grounded for an individual, it must also be grounded for others in order for conventionalization to occur. The details of this process are similar but not identical to the initial reinterpretation of novel symbols. Again, communicative intent must be recognized while not perceiving nonarbitrary features in a sign, and intended meaning is inferred from other factors. Unlike initial reinterpretation, the possibility for retrieving meaning from observation of others' successful use of nonarbitrary signs is not available. Inference from contextual information alone allows new symbols to be correctly interpreted and adopted by others, which requires that interlocutors can otherwise maintain a shared focus of attention. Thus, while contextual inference based on joint attention is one possible route for reinterpretation, it is crucial to the subsequent spread of novel symbols. This suggests the conventionalization process requires a rich and supportive communicative superstructure in which novel arbitrary signs could be used, with overall comprehension and shared understanding maintained so that intended form-meaning relationships could be correctly interpreted.

It is of course true that observation could again serve as a grounding mechanism once multiple individuals share symbols. Indeed, it is likely an instrumental amplifier in the conventionalization process; as more people share a symbol and communicate with it, the more opportunities there are for observational transmission to others, therefore increasing the speed at which the symbol spreads throughout the remaining population. However, how symbols become shared by more than one individual to begin with must first be explained before a possible role for observation could come into play.

6. Conclusion

Research reviewed in this chapter has illuminated key factors for emerging symbolic communication in the abstract. In addressing the issue in the case of human evolution, it is also necessary to consider the particular characteristics and circumstances of human communication. Medium of expression determines representational capacity of signs, and by extension influences how signs are perceived with regards to nonarbitrariness and arbitrariness. Therefore, we must also examine the modalities through which humans communicate and the way each is used to convey meaning. The following chapter will explore the multimodal nature of communication and discuss its relation to the evolution of symbols.

CHAPTER 3

Human symbolic communication and emergence

1. Introduction

Preceding chapters have helped to frame the question of symbol origins in the abstract, in terms of theory and general processes. A further important factor that influences how symbols arise is the modality or modalities through which communication takes place, and this element will vary depending on the specifics of the communicators and their situated experience. Therefore, to address the specific question of how symbolic communication emerged in human evolution, it is essential to consider the particular features of human communication – the anatomical and cognitive attributes of the communicators themselves. This chapter will first discuss these issues and situate the present inquiry in an embodied perspective, which will serve to further frame our question and construct the explanatory landscape from which an answer can be derived.

Human communication is not limited to symbolic speech, and a substantial amount of information is expressed through other semiotic strategies in the vocal channel as well as bodily movements. Some of these additional forms of communication are also referential, if not symbolic. As such, it is important to understand how meaning is represented in these other ways, and in addition, if and how it is related to symbolic communication. The first two sections of this chapter will be devoted to exploring the two major modalities of human communication: vocal and gestural. Topics addressed include how each is utilized to convey meaning, how each operates in language and how the two interact anatomically and semiotically. Following that, conclusions regarding the multimodal nature of communication will be discussed in the context of human evolutionary history in order to characterize the communicative conditions from which symbols would have emerged. Finally, in order to better understand how nonarbitrary signs in each modality are perceived and interpreted, their semiotic profiles will be compared based on physical qualities and the constraints those impose.

With these ideas in place, the remainder of the chapter will directly address the question of symbol emergence in human evolution, including a critique of existing theories and solutions made available from the ideas and evidence presented in this and preceding chapters.

2. The vocal modality

The traditional view of language is that the words of speech are symbols and arbitrarily refer to their meanings (Hocket, 1960; Saussure, 1959). Symbolic communication thus conceived is the expression of meaning through arbitrary sounds. As the characterization of speech as predominantly symbolic is uncontroversial and widely acknowledged, this aspect of vocal communication will be accepted as a given and not reviewed in detail. However, non-symbolic aspects of vocal communication, which typically are not emphasized in discussions of language, warrant further review.

While vocalization is no doubt the primary domain of symbolic communication in normal circumstances, the modality potentially can represent via nonarbitrary semiotic strategies, and some argue this is a general and pervasive quality of spoken language (Perniss, Thompson & Vigliocco, 2010). New research has begun to explore precisely how nonarbitrariness works in the vocal channel and the extent to which it exists in the words of speech.

Vocalizations are auditory and have a temporal aspect, and can therefore resemble sounds and sonic events. Vocal iconicity is commonly known as onomatopoeia, examples of which are English words like *crackle* or *buzz* that are meant to mimic certain noises. Vocal iconicity appears to be limited, however, as different languages often use completely different sounds to mimic the same noises. For example, in English the sound a pig makes is pronounced /oɪŋk/, while in Japanese it is /bu:/. This inconsistency suggests the iconic connection between vocal sounds and their referents is somewhat weak and conventionalization is also a major factor in their interpretation and use.

Although auditory resemblance through vocalization is somewhat limited, speech sounds can also be used to nonarbitrarily represent in a way that is not directly iconic. By way of crossmodal associations, sonic forms can resemble forms in non-auditory modalities (Nuckolls, 1999; Simner, Cuskley & Kirby, 2010). These crossmodal associations are based on relating sounds to experiences in vision, touch, taste and olfaction, and these are often used to depict movement, size, shape,
color, and texture (Hinton, Nichols & Ohala, 1994). One example of words that appear to derive from crossmodal associations is what are known as *mimetics* in Japanese, which refer to psychological or physiological states, manner of motion, the temporal structure of events, and affective content associated with events and states (Kita, 2008). For instance, the word *koro* is used to describe a light object rolling while the word *goro* describes a heavy object rolling – that is, the higher frequency, voiceless /k/ sound is associated with light-weightedness while the lower frequency, voiced /g/ sound is associated with heaviness.[7] Finally, in addition to speech sounds, suprasegmental features of vocal language like prosody can also convey information nonarbitrarily, like intensity and size. For example, pronouncing the word *big* with a low voice and extending the syllable beyond the expected rhythm in order to convey largeness.

While the specific cognitive mechanisms responsible for crossmodal associational abilities are not yet well understood (Ahlner & Zlatev, 2010), a growing body of research suggests speakers naturally use these abilities to map forms to meanings (Perniss et al., 2010). Moreover, multiple studies have demonstrated that some crossmodal biases are shared across cultures (reviewed in Cuskley & Kirby, in press), indicating that specific sound-to-other-modality associations may be universal. For instance, the pitch-to-size association cited above has been found across cultures (Kita, 2008). Shared biases of this sort provide an opportunity for resemblance between forms and meanings in different modalities to be perceived and used for

[7] Although mimetic words appear to be nonarbitrary via crossmodal associations in this way, they differ from typical nonarbitrary gestures (discussed below) in that they are fully conventionalized items in the Japanese vocabulary. This conventionalization likely exists because mimetic words largely exhibit secondary iconicity (a concept developed by Sonesson, 1997 and discussed in Chapter 1), which means their form alone would generally be insufficient to convey meaning without learning the specific association beforehand.

communicative purposes. Thus, the vocal modality has the potential for nonarbitrariness beyond direct iconicity. Kita (2008) notes, however, that the types of events and states that can be referred to through crossmodal iconicity are somewhat restricted.

Ahlner & Zlatev (2010) characterize nonarbitrariness in spoken communication using the distinction between primary and secondary iconicity mentioned in Chapter 1 (Sonesson, 1997). Primary iconicity, in the terms of this thesis, is when a sign has high transparency in form (Chapter 2, section 2.2.2) – when meaning is sufficiently interpretable from recognition of features depicted in a form. Secondary iconicity is when correspondences between form and meaning are only perceived *after* becoming aware of a sign relationship. Once the referent is known, form-meaning similarities that would not otherwise be perceived immediately become apparent. Ahlner & Zlatev (2010) review research on cross-cultural comprehension of mimetic words and note successful interpretation only occurs when two preconditions are met. First, participants are presented with a list of familiar words in their own language to match with unfamiliar words, giving them prior knowledge of specific meanings and allowing them to search for features of those meanings in the list of unfamiliar words. Second, the list of familiar words must be made up of antonym pairs, which highlights salient distinctions in meanings and further aids in the search for form-meaning correspondences. Ahlner & Zlatev argue these results indicate that much of what is considered sound symbolic in spoken language in these experiments does not actually rely on perceptions of nonarbitrariness as the dominant strategy for interpretation. They go on to note that, as novel signs generally start out as iconic, such words may have shown greater primary iconicity in the past when they served to establish communication. But because these words become conventionalized, any pressures to retain iconic features are eased, thus allowing their forms to change through learning and transmission processes.

2.1 Discussion

The evidence shows that, in addition to the traditionally recognized function of conveying arbitrary symbols, the vocal modality can be used for nonarbitrary representation. Furthermore, its representational capacity is expanded by the ability to

perceive resemblance crossmodally between vocal forms and other non-auditory sensory experiences. Thus, in addition to communication through conventional symbolic words, vocal signs can be created and interpreted based on shared sensory knowledge and shared perceptual biases.

While much of speech presumably is perceived as arbitrary and understood as symbolic by speakers, it is of course possible that, separately from known mimetic words, some nonarbitrariness is perceived in some words by some individual speakers. However, these are likely idiosyncratic and it is highly improbable that any particular perception is held by a large portion of the population, and therefore not utilized as a communicative device. Taking into account these possible exceptions, we can nevertheless accept that, proportionally and on a population-level of analysis, the majority of speech is interpreted as symbolic.

Interestingly, although the vocal modality can convey meaning in these non-symbolic ways, nonarbitrary vocalizations are highly conventionalized and incorporated into the vocabulary of a language. One reason for this lexicalization may be in part because most vocal signs are symbolic and conventionalized by necessity, which could instill an expectation or bias for conventional vocal forms. Another, perhaps more important reason is that nonarbitrary words show low transparency in form, which means additional information is required initially to connect form to meaning. As Ahlner and Zlatev (2010) state, earlier forms of these modern words may have carried more nonarbitrary features, which would boost their communicative capacity somewhat and perhaps enable interpretation in the absence of conventional knowledge.

While it may be that nonarbitrary vocalization has greater potential for referential capacity than the examples found in modern language, we can nevertheless note that the physical qualities of the modality itself place substantial constraints on this capacity, particularly in regard to human perception and action. Human interaction with the world is dominated by rich and detailed visual-spatial experience. Even with the aid of crossmodal associations, sounds can resemble aspects of this experience to only a limited degree. For instance, vowel quality can evoke a very general idea of shape, such as roundness versus sharpness. This information, on its own, only hints at

the larger meaning of a message. More fine-grained features of a specific intended referent must be identified for correct interpretation to occur, which requires additional information from some other communicative resource (see Chapter 2, sections 2.2.1 and 4.1 for what these might be). This argument is not to suggest that nonarbitrary vocal communication is trivial or irrelevant to human communication. The nonarbitrary information it does carry enriches and likely eases communication, and although the current discussion is focused on referential communication of semantic content, the importance of affective information often carried in vocalization should not be underestimated. Concerning the capacity to convey sufficiently numerous and specific features of a referent for effective comprehension, however, the vocal channel is inherently restricted. Of course, a more informative approach to the issue of representational capacity would be to take a relative perspective and consider how the vocal modality compares to other representational channels, which will be addressed in section 5 after the gestural modality has been examined in turn.

One final point to address is the idea of non-conventional nonarbitrary vocal communication. Mimetic words are not the only way vocalization is nonarbitrary. As mentioned above, stress and intonation patterns overlaying speech can convey some nonarbitrary features. Spontaneously produced, idiosyncratic nonarbitrary vocalizations have not been studied or documented to the degree that gestures have, though it is quite easy to think of examples from everyday communication - for instance, to embellish storytelling. In addition, iconic or cross-modally iconic vocalizations are often produced when 'acting out' an idea, like when a child pretends to be an airplane and makes noises associated with planes at the same time as using their body to resemble the shape and movement a plane. At the very least, the ability to create nonarbitrary vocalizations is not in doubt. Such vocalization pervades natural communication yet, like other forms of nonarbitrary communication (see following section on gesture), are typically processed at a level below conscious awareness.

3. Gesture

Until relatively recently, gestural communication was a relatively under-researched topic. Following the seminal works of Kendon (1972) and McNeill (1994), increasing

interest in gesture has resulted in an abundance of new research. Technological advancements that allowed for detailed analysis of hand, arm, facial and other bodily movements aided this research, and it has become clear that gesture, in the broadest sense, serves multiple important functions in natural communication.

The term *gesture* can encompass a wide assortment of communicative body movements that differ from each other in important ways. Therefore, it is first necessary to limit the scope of discussion and identify the type or types of gestures that are the relevant to this thesis and the present examination of human symbolic communication specifically.

3.1 Delimiting gesture

Bodily movements of various types can be categorized by i) the kind of information they convey, ii) their semiotic status and iii) when and how they are produced in relation to other communicative processes. Regarding i), some gestures are related to emotion and express mood, tone, intensity or other affective information (Ekman & Frieson, 1969; Maricchiolo, Gnisci, Bonaiuto & Ficca 2008), while others are representational and carry semantic content (McNeill, 1992). It is this second category – gestures that function as signs – that are of interest here.

Within the set of representational gestures, further distinctions can be made along dimensions ii) and iii). Regarding ii), these gestures differ in the type of sign they embody – whether they are nonarbitrary or symbolic. McNeill (2000, 2005) also divides them according to point iii), whether they are produced in accompaniment to speech. A number of classification systems have been proposed (Burling, 2005; Chawla & Krauss, 1994; Efron, 1972; Hadar, 1987; Hadar & Pinchas-Zamir, 2004; McNeill, 1994, 2005; Rauscher et al., 1996), but perhaps due to the relatively short history of gesture studies, no single one has been adopted universally. Because McNeill's (1994, 2005) system is widely accepted and provides the most comprehensive and clear categorization of gesture types based on both function and form, it will be used for this discussion. Under McNeill's system, there are four types of gesture: gesticulations, pantomime, emblems and signs.

Gesticulations are nonarbitrary gestures that are produced while speaking. They can be iconic and/or indexical, like movements that portray actions and entities or pointing to refer by way of spatial location. Gesticulations are grouped into *iconics*, *metaphorics* (also iconic, but these depict abstract concepts or relations) and *deictics*. There are also certain gesticulations called *beats* that are more abstract and related to the rhythm of speech and discourse structure. They are simple in shape and movement, consisting of small flicks or oscillating actions and generally reflect how the accompanying speech fits into a larger discourse organization.

McNeill (2005) has since claimed that the semantic content contained in a given gesticulation is of much greater theoretical interest than its classification, as during natural gesture realization these four subcategories are not mutually exclusive. For example, beats can be superimposed onto any of the other categories, and an iconic gesture can simultaneous express deixis if the body part portraying the iconic shape is moved in a direction to indicate the location of the referent.

Pantomime is the production of nonarbitrary gestures in the absence of speech and is used when the auditory channel is compromised or unavailable, such as in a loud factory or in play. Typical examples of pantomime are the movements produced in the game of charades, in which players 'act out' an idea in silence for others to guess using iconic and/or indexical gestures.

Emblems are described as gestures that seem to be arbitrarily related to their meanings, and, as a result, are conventionalized. An example is the action of pointing to one's temple and rotating the forefinger in a circular motion, which means something along the lines of 'crazy'. Emblems typically occur with speech but are also used on their own.

Signs are gestures that constitute words and/or grammatical functions in sign languages and exhibit all the same properties of spoken language, most notably a greater proportion of conventionalized and arbitrary signs.

According to McNeill, gesticulations have a unique relationship to speech not found in other gesture types. Gesticulations are closely tied to the symbolic code of

language – that is, they co-occur in tight temporal and semantic coordination with another form of meaning expression. Moreover, a gesticulation typically must be tied to the utterance with which it occurs in order for the full meaning to be interpretable. To illustrate this obligatory conjunction with speech, McNeill (2005) describes a speaker who gestured an arc shape with his hand while saying 'he grabs a big o[ak tree and he **bends it way back**]' (pg. 6, coding original):

> His hand rose from the armrest of the chair as he said "oak" (left bracket), reached its apex with "he", at which moment there was a brief prestroke hold (underlining); the hand then moved downward and to the side during the boldface section (the stroke – the part of the gesture depicting the actual 'bending back')… At this point there was a poststroke hold and a new gesture began.
>
> During the stroke phase, the hand appeared to grasp and bend back an object with some thickness. Such a gesture has clear iconicity – the movement and the handgrip; also a locus (starting high and ending low) – all creating imagery that is analogous to the event being described in speech at the same time (a comic book character bending back an oak tree).
>
> …The bends-it-back gesture is meaningful only in conjunction with the utterance of "bends it back." (pgs. 6-7)

Evidence to support this claim of necessary coordination comes from a study by Hadar & Pinchas-Zamir (2004), which demonstrated that, although at times it may be possible to infer some very general meaning from gesticulations isolated from speech, without access to concurrently spoken words, the *specific* meaning that a gesture represents is unrecoverable.

In contrast to gesticulations, other gesture types do not require any additional information for their meanings to be understood. Emblems have conventionalized meanings and can convey that information without the aid of speech. As pantomime does not occur with speech by definition, it must depict referents through more elaborate nonarbitrary representation than is normally found in gesticulations. Although gestural sign languages and vocal speech utilize separate modalities, and

therefore co-production is theoretically possible, signers find it very difficult to use speech and signs simultaneously. A later study by Fontana (2008) found evidence for this disharmony of speech and sign by showing that the performance of both systems suffers when participants attempted to produce both at the same time.

McNeill (2005) goes on to contrast gesticulations and other gesture types with regards to resemblance to a linguistic system, level of conventionalization and representational characteristics. These contrasts are used to highlight how gesticulations exhibit qualities that are wholly distinct from and complementary to those of signs and speech. The details of his proposal will not be reviewed here, but the overall argument is that these qualities are a consequence of the different semiotic status of gesture types and their relationship to speech.

3.1.1 Discussion

Put in the terms introduced in Chapter 1 in the discussion of semiotics, one aspect of McNeill's categorization is to separate nonarbitrary and arbitrary representations. Gesticulations and pantomime are nonarbitrary, and this characterization is fairly uncontroversial. Emblems are said by McNeill to be arbitrary, though their semiotic status is arguably less clear-cut. While emblems are not as obviously iconic, it may be that some nonarbitrariness is perceived in their forms. For instance, the example of the gesture for 'crazy' given above in some ways depicts features related to the meaning; it is head-oriented, and craziness is thought to reside in the brain; the circular movement of the finger implies something along the lines of disorder, which is another concept associated with insanity. The fact that emblems are conventionalized could help to explain why their semiotic status is less certain, as nonarbitrary features are not the sole carrier of meaning when shared cultural knowledge of form-meaning mapping is available. In this way, emblems may be thought of as largely exhibiting secondary iconicity (Sonesson, 1997; Chapter 1, section 3.2.1). Finally, signs are more straightforwardly arbitrary. Of course, not *all* signs are arbitrary (Chapter 2, section 3), but as arbitrary gestures are found in sign languages with much greater frequency than in any other setting, they are a useful prototype for this category. Thus, the gestural modality, in various circumstances, is employed for every kind of representational strategy – iconicity, indexicality and symbolism.

Another primary distinction between McNeill's categories is relationship to speech. Manual signs (in McNeill's terms) are not used frequently by hearing people while speaking, most obviously because the vocal modality is already occupied with symbol production. Emblems are used both with and without speech, and this less dichotomous relationship mirrors the uncertainty of their semiotic status. In contrast to the incompatibility of speech and arbitrary gestures observed in bilingual signers, gesticulations are easily, naturally and frequently produced at the same time as speech. Finally, pantomime is defined by an absence of speech and typically only occurs when the vocal modality is restricted in some way. Taking into account a gesture category's relationship to speech in this way helps in determining which gesture type or types are most relevant to the emergence of symbolic communication in human evolution, and each of these cases warrants further discussion.

First, the fact that manual symbolic signs do not occur with speech is in large part due to their being a non-normal mode of communication. Enculturation in modern language that takes place during typical human development renders vocalization the default and predominant vehicle of symbolic communication. The gestural modality overwhelmingly is only enlisted for this function when auditory channels are compromised.[8] More tellingly, when signing and speaking are attempted at the same time by bilingual hearing individuals, production is disrupted in both modalities. Thus, bodily symbolic signs are disassociated with vocal symbolic communication not only as a result of culture and development, but also due to an apparent fundamental incompatibility at the level of simultaneous online processing. Even though gesture appears to be an integral component of human communication, symbolic gestures like those of sign language are not an integral part of symbolic communication as this thesis seeks to understand it – as it occurs in typical and natural conditions. The existence of bodily symbolic signs and the conditions in which they do occur are still important clues for an investigation of symbolic communication and are considered elsewhere (Chapter 2 and in section 6 below). However, in identifying the unique characteristics of typical human communication

[8] Some isolated exceptions are manual systems that have been developed in response to culture-specific taboos on using spoken language to communicate with certain people and/or about certain topics (Kwek & Kendon, 1991).

generally, around which we can formulate an explanation of symbol emergence, bodily symbolic signs will not be a central focus.

Emblems are again a somewhat problematic case. Although emblems can be produced while speaking, they can also be – and often are – used as a substitute for speech. If emblems are in fact sometimes arbitrary, perhaps this partial disassociation is due to a difficulty in processing like that between signed and spoken language. Alternatively, it may be that these highly conventionalized gestures carry the bulk of a desired message and speakers employ them unimodally simply for efficiency. Given the human propensity for imitation, it is not surprising that conventionalized gestures – arbitrary or nonarbitrary – exist. And this same phenomenon was also likely in action in the stages of human evolution of interest here. However, as emblems do not show an obligatory and intimate connection to the primary mode of symbolic expression, we will for now consider them a minor feature.

By definition pantomime is fully disassociated from speech. While it is unrelated to symbolic communication this way in an online sense, nonarbitrary gestural communication like pantomime is still relevant to the question of symbol origins and will be addressed at a later point. For now, one issue regarding the distinction between pantomime and gesticulations should be mentioned. In other writings, Singleton, Goldin-Meadow & McNeill (1995) have argued that the two stem from quite different cognitive systems. Nonarbitrary gestures do take on different qualities when people are asked to mime short events as compared to their co-speech gesticulations in describing the same event in words, notably becoming more strictly ordered and analytic – that is, more speech-like (Goldin-Meadow, 2008). However, gesticulations and pantomime share many features, arguably more fundamental ones than those distinguishing them. Both are explicitly communicative representations – *signs*, in the terminology of Chapter 1 – and both are clearly based on sensorimotor knowledge of actions and objects. From the perspective of hierarchical models of cognition (A. Clark, 1997; Damasio & Damasio, 1994) the neural systems underlying production and comprehension of each likely overlap a great deal, and as both involve voluntary, referential communication, this overlap would extend beyond low-level sensory systems. Differences between the two in form and execution seem attributable to matters of attention and context, and thus do not warrant distinct

cognitive explanations. Given that pantomime embodies many semiotic qualities of interest to this thesis, specific questions of how it relates to gesticulations, vocal communication and the emergence of symbols will be considered later in this chapter and in following chapters.

Finally, gesticulations only occur with speech and nonarbitrarily represent meaning directly related to the meaning concurrently expressed in words. As such, this type of gesture is directly relevant to characterizing human symbolic communication accurately. The fact that the manual modality is utilized as a complementary semiotic strategy to express simultaneously part of a message also conveyed in vocal symbols suggests the nature of language may be more complex than traditional conceptions of symbolic communication as speech in isolation.

To better understand the particularities of human communication regarding these issues it is necessary to further explore gesticulations and how the gestural and vocal modalities interact during the expression of symbols. The following section will provide a brief overview of research on gesture and language and discuss its implications for the questions at hand.

3.2 Speech and gesture

This section will present research on the relationship between speech and gesture, examining the anatomical and semantic links between the vocal and gestural modalities; how gesture affects cognition and communication; and finally gesture and language in ontogeny. For purposes of simplicity, the term *gesture* will be used for the remainder of this section to refer to co-speech gesticulations unless otherwise explicitly stated.

3.2.1 Semantic content and co-expression

McNeill (2005) describes how gesture is co-expressive but non-redundant with speech. It is co-expressive in that the meaning embodied in a gesture is linked semantically and temporally with the corresponding meaning conveyed in speech (McNeill, 1994). However, gestures do not carry only meanings explicitly expressed in words, and oftentimes gestures are used to represent aspects of an overall message

that are in fact absent in speech. In other words, the two modalities jointly express an underlying *idea*, some features of which may be expressed in speech only, in gesture only or in both modalities at the same time. In many cases gesture alone carries meaning that is crucial to conveying an intended message. This has been demonstrated experimentally by showing that speakers express in gesture information required to understand an intended message while omitting it from speech (Melinger & Levelt, 2005), with one study finding that 90% of gestures produced in face-to-face conversations conveyed some crucial information not included in the accompanying speech (Bavelas, Gerwing, Sutton & Prevost, 2007). In addition, if common ground is established (H.H. Clark, 1996; Chapter 1, section 3.2.2) – a feature of communication in natural settings – speakers use gesture to convey an even greater proportion of content (Holler & Wilkin, 2008).

Examining gesture production cross-linguistically has helped to shed light on the interaction of speech and gesture, as different languages have different ways of expressing semantic features lexically, morphologically and syntactically (Kita, 2009). Kita and colleagues (Kita & Ozyürek, 2003; Kita, Ozyürek, Allen, Brown, Furman, & Ishizuka, 2007) have found that speakers represent an event in gesture in a similar way to its representation in speech, which means that gestures are influenced not only by the images on which they are based, but also by the specific structure of the coordinated speech. The connection between linguistic structures and gesture execution has been further illustrated in a study of an aphasic patient (Kemmerer, Chandrasekaran & Tranel, 2007), whose gestures produced while attempting to speak exhibited features of English verb and preposition constructions.

Wagner, Nusbaum & Goldin-Meadow (2004) tested the representational status of co-speech gestures and conclude they constitute propositional representations with semantic content, not merely visual-spatial sensorimotor representations. Similar findings in neuroscience indicate that co-speech gestures are enlisted specifically for semantic processing, and their meaning is used to reduce semantic ambiguity (Holle & Gunter, 2007; Kelly, Kravitz & Hopkins, 2004; Ozyurek, Willems, Kita & Hagoort, 2007; Skipper, Goldin-Meadow, Nusbaum & Small, 2007; Wu & Coulson, 2007; Xu et al., 2009). In line with these findings, when processing utterances listeners do not discriminate between vocal and manual sources, incorporating information conveyed

in gesture into their understanding of a verbal expression (Kelly, Barr, Church & Lynch, 1999; Kelly, Ozyürek & Maris, 2010; McNeill, Cassell & McCullough, 1994). From such evidence, McNeill et al. (1994) argue that the speech and gesture channels 'smoothly combine into a single idea unit' (pg. 235) and propose that both production and comprehension in linguistic communication involve the integration of the two modalities.

3.2.2 Cognitive and communicative benefits

The act of gesturing affects memory and other aspects of cognition for both speakers and listeners in intriguing and what may at first appear surprising ways. Various methods of restricting movement and thereby preventing gesturing have been shown to cause speech dysfluencies (Morsella & Krauss, 2004; Rauscher, Krauss & Chen, 1996; Rime, Schiaratura, Hupet & Ghysselinckx, 1984), and another study demonstrated this effect is not simply a byproduct of overt restriction or memory demands that could arise when participants are asked not to gesture (Alibali, Heath & Myers, 2001). Other research has found that producing simultaneous gestures enhances children's performance in picture naming tasks (Pine, Bird & Kirck, 2007), and similar effects have been found in aphasics (Hanlon, Brown & Gerstman, 1990). In addition, the number of gestures people produce while speaking increases as the cognitive demands of a task increase, indicating doing so aids performance on the task in some way (Alibali Kita & Young, 2000; Hostetter, Alibali & Kita, 2006; Kita & Davies, 2009; Morsella & Krauss, 2004; Wesp, Hesse, & Keutmann, 2001). Gesture can also aid speakers' memory for information not related to the content of speech, as gesturing during an unrelated task was found to increase memory for information encountered in a previous task (Goldin-Meadow, 2003b; Wagner, Nusbaum & Goldin-Meadow, 2004).

Some researchers contend the role of gesture in speech production is facilitative, (Iverson & Goldin-Meadow, 2000; Krauss, Chen & Chawla, 1996; Rauscher et al., 1996), meaning that gesturing facilitates access to symbolic words while speaking. However, others argue based on research on stuttered speech that this may not necessarily be the case. Mayberry & Jacques (2000) examined the gestures produced during periods of stuttering and found that these do not appear to be compensating for the degraded content and structure of speech. Instead, fewer, not more, gestures

accompanied stuttered speech compared with fluent speech. In addition, gesture execution was precisely temporally aligned with speech execution despite the many interruptions caused by stuttering. The authors interpret these results as indicating that the coordination of speech and gesture takes place before either is expressed in production. If so, this would mean gesture is not primary to and enabling the retrieval of spoken words but is instead accessed and produced in precise alignment with speech.

Gesture has been shown to have beneficial memory effects for hearers as well as speakers. Recall of speech is significantly better for speech-plus-gesture utterances than speech alone, and this holds at the word (Kelly et al., 1999; Kelly, McDevitt & Esch, 2009), sentence (Feyerseisen, 2006) and discourse level (Beattie & Shovelton, 1999). In addition, judgments regarding the intention of indirect speech acts are significantly more accurate with speech-plus-gesture utterances than both speech-only and gesture-only combined (Kelly et al., 1999).

3.2.3 Ontogeny

Developmental research demonstrates evidence for initial and emerging connections between the vocal modality, manual modality and language. A phenomenon known as the Babkin reflex, wherein newborns respond to pressure applied to their palms by opening their mouths, illustrates the early and direct anatomical links between oral and manual modalities (Iverson & Thelen, 1999). 'Manual babbling' observed in infants co-occurs with vocal babbling, and such manual actions do not also correlate with other motor milestones (Bates & Dick, 2001). Iverson & Thelen (1999) cite these initial biases for hand-mouth coordination as establishing a firm connection that becomes increasingly elaborated as language development progresses.

Bates & Dick (2001) review research on the word stage of language development. At around eight to ten months of age children begin to show evidence of word comprehension, and this is correlated with the first uses of deictic gestures. Comprehension and production of pantomime are both correlated with auditory language comprehension (Bates & Dick, 2001), indicating that the interpretation of meaningful gesture is in some way fundamentally related to the interpretation of speech. This connection is made more apparent by the delay of both word

comprehension and early gestural production in 'late talkers' as well as children with certain brain injuries, signaling that the understanding of symbolic representation and first expression of nonarbitrary representation emerge simultaneously in the vocal and manual modalities.

Other studies have examined precisely how children coordinate the vocal and manual modalities in their early utterances. Children's early communicative acts are typically combinations of deictic gestures and vocalizations (Bates & Dick, 2001; Capirci & Volterra, 2008; Pizzuto & Capobianco, 2005). The most frequent utterances made by children have been found to be deictic gestures plus spoken words (Capirci & Volterra, 2008; Iverson & Goldin-Meadow, 2000). Capirci & Volterra (2008) found that meanings expressed in children's early gestures do not necessarily precede the same meanings being expressed vocally, as expression in both could appear at the same or different times.

Children's use of gesture in intermediate stages of language development has not been studied as extensively as earlier stages. However, one of the few studies undertaken investigated how speech and gesture develop in the different languages of bilingual children (Mayberry & Nicoladis, 2000). The authors found that the emergence of iconic and beat gestures coincided with the onset of longer, sentence-like utterances, and this occurred differentially in the separate languages. For example, if children were at the multi-word stage in one language, they would produce iconic and beat gestures with these longer utterances; however, if they remained at the one-word stage in the other language, they never produced the same types of gesture when speaking. Mayberry & Nicoladis (2000) argue these findings demonstrate that gesture development is not independent from, but actually corresponds to, language development.

The course of language and gesture development also sheds light on the nature of meaning representation and expression. In the case of non-normal development, children with Specific Language Impairment (SLI) perform poorly on tests of nonarbitrary gesture imitation, even while some perform at normal levels on other movement tests (Iverson & Thelen, 1999). Deloach (1995) found that understanding of iconic representation occurs only after language development is under way. A

more recent experiment designed to test children's recognition of iconicity showed that this skill does not reliably emerge until around 26 months – at the same time as or after the development of some verbal skills (Namy, 2008). Namy contends her results challenge the traditional view that iconicity facilitates the use of symbols and suggests that a more general understanding of referential representations may underlie the recognition of both iconic and symbolic representations. This claim is supported by another study that found children's ability to map iconic numerical signs (fingers) onto various quantities of toys showed no advantage over their ability to do so with arbitrary signs (Arabic number symbols), with performance on arbitrary signs reaching even slightly higher than iconic ones (Nicoladis, Pika & Marentette, 2010).

3.2.4 Gesture as language

Based on the wealth of research demonstrating how the meaning physically expressed in gesture is done so synergistically in conjunction with speech, some authors claim that gesture is not merely a by-product or accompaniment *to* language, but is in fact an intrinsic component *of* language. McNeill (2005) contends that language is fundamentally multimodal, and the two modalities in coordination constitute a unified speech-gesture system. Bates & Dick (2001) argue that the coupled developmental trajectory of vocal and gestural communication demonstrates that the cognitive systems underlying the perception and spontaneous production of speech are shared by those for the perception and spontaneous production of gesture. Along the same lines, Iverson & Thelen (1999) note that very early sensorimotor biases linking oral and manual modalities initiate a deep connection and eventually 'cascade into a single, coupled, communicative system, where the mental aspects of expression are manifest in movement' (pg. 36). Similarly, according to Gallagher (2005), gesture and speech are part of a common psychological structure, forming a complementary, coupled system of meaning expression. In other words, on this view, gesture is not related to language, gesture *is* language.

Somewhat more anomalous examples provide evidence to support the position that nonarbitrary co-speech gestures are not merely communicative, but specifically linguistic. Congenitally blind speakers, who have never observed gesturing themselves, do so in a virtually identical manner to sighted speakers, even when their addressee is also blind (Iverson & Goldin-Meadow, 1998). Anecdotal evidence also

comes from patients with missing limbs from birth who claim to experience 'phantom gesturing' while they speak (Ramachandran & Blakeslee, 1998).

A remarkable pathology case study conducted by Cole, Gallagher & McNeill (2002) further informs this issue. Due to a previous injury, the patient lacked all proprioceptive abilities and could not direct even the most basic bodily movements without conscious, visual monitoring. However, when visual access to his hands and arms was blocked experimentally, he continued to produce unconscious, unmonitored co-speech gestures, and these remained consistent with normal measurements in timing and shape. The authors argue these findings indicate that linguistic meaning itself – not simply the motor system – is responsible for gesture production.

3.2.5 Discussion

The evidence presented here has revealed a deep connection between vocalization and bodily gesture, in parallel with an equally strong relationship between symbols and nonarbitrary representations as they are expressed in the corresponding modalities. Speakers coordinate the two modalities spontaneously, utilizing complementary semiotic strategies to communicate through both.

The reviewed research demonstrates that nonarbitrary gesture is intertwined with symbolic expression, being intimately involved in processes of both speech production and comprehension. It is influenced not only by the communicative context of speaking, but also by the specific structures of a speaker's linguistic system. Kita and Ozyurek's (2003) cross-linguistic study shows that the way meaning is manifested in gesture is directly related to the way meaning is conceived and expressed in speech. This suggests gestures are not simply a product of spatio-motor control, but are instead part of the linguistic system, constrained and shaped by its grammatical and semantic properties. These results also provide insights into the nature of gestural meaning. Participants, regardless of language, regularly encoded direction in their gestures while never including this information in speech. Direction, in this case, was an essential component of the speakers' intended message. That such a necessary feature was expressed in gesture indicates it is a critical component of language and derives from the underlying idea a speaker is attempting to convey through words.

The benefits associated with co-speech gesture also highlight its linguistic nature. The communicative benefits gained by hearers is not unexpected, as gestures carry meaningful information related to speech and not utilizing this information would be a sub-optimal strategy for comprehension. The cognitive benefits for speakers, however, are somewhat more difficult to interpret. As noted above, many authors contend that gestures are playing a facilitative role in word (or larger linguistic unit) retrieval or conceptualization processes for speaking (Alibali et al., 2000; Hostetter et al., 2006; Kita & Davies, 2009; Pine et al., 2007; Rauscher et al., 1996). The characterization of facilitative would seem to imply that gestural processes are *primary* to those of speech, involving some sort of priming or feedback system. However, the full data set presented indicates that the two processes are simultaneous; gesture is not facilitating the expression of meaning in speech, but instead is concurrently expressing an underlying meaning that may or may not be verbalized. Because speech and gesture representational systems are shared (Chapter 1, section 4), this activation is likely simultaneous, which suggests gesture is not an *aid* for finding words but a parallel output that is a natural *consequence* of expressing meaning in words. If the data are interpreted to mean that human communication is, at its core, multimodal – that intended meaning is expressed through the complementary channels of vocalization and bodily movements – a much simpler explanation is available for the observed cognitive effects. Speech dysfluencies from gesture restriction – both experimentally- and self-imposed – may not be signaling gesture's role in accessing linguistic units and could merely be the result of inhibiting the most natural form of expression. Reduced memory for incidental information could also be explained by this account. If speakers are not enlisting both modalities and distributing the cognitive burden of expression between them, other, more general cognitive capacities could suffer as a consequence of placing the full computational load onto the symbolic channel.

Research on ontogeny shows that the anatomical and semantic coordination between the modalities is observable from the earliest stages of development. Children's ability to express meaning – to represent for communicative purposes – is not only multimodal and semiotically complementary, but that ability develops in perfect synchrony across the two modalities. Meaning is initially represented in each

modality contemporaneously, and children immediately begin to *co-express* the same or combinations of meanings in both modalities *simultaneously*. Importantly, gesture is not an alternative form of expression that children rely on to enhance their communication before verbal skills are sufficient. Instead, gesture development starts and progresses along with vocal development; children's earliest communication consists of multimodal, heterosemiotic utterances, and utterances remain multimodal as their language skills become more sophisticated. Overall, developmental research corroborates evidence from adult studies for a unified, multimodal language system and suggests a joint representational foundation for symbolic speech and nonarbitrary gestures.

The developmental parity of nonarbitrary and symbolic signs has another important implication. The fact that the ability to understand and use both sign types arises together indicates that the capacity and inclination to interpret symbols is not a separate or higher cognitive phenomenon to interpreting nonarbitrary signs. Instead, the capacity for intentional, referential representation appears to encompass all three representational strategies in one package. A similar phenomenon is also displayed in children's ability and inclination to imitate arbitrary behaviors in non-communicative practical tasks. Horner & Whiten (2005) conducted an experiment wherein children watched an adult perform a task through a series of actions, some of which were not necessary to reach the desired goal. Despite the fact that it would have been apparent from the children's perspective that certain actions were unnecessary, they nevertheless copied the full series of both required and irrelevant behaviors. These results, though not related to communication specifically, do illustrate that the skill and proclivity for learning and imitating arbitrary behaviors is a basic feature of the human imitative abilities that ultimately underlie and enable bodily mimesis, which further supports the idea that arbitrary signs would be learned and used in a similar way. Thus, with this combined evidence in mind, we can reason that when the ability to communicate through nonarbitrary signs arose in human history, it also entailed the *potential* to use symbols.

The key conclusion to draw from the review of gesture and speech is that human referential communication is fundamentally multimodal, heterosemiotic and coordinated. That is, it appears the natural mode of linguistic expression is to i)

combine nonarbitrary and symbolic sign types; ii) differentially distribute those sign types between the vocal and gestural modalities; and iii) simultaneously produce semantically related signs across modalities.

One point to note is that the assertion that nonarbitrary gesticulations are not primary to symbolic speech is specifically in regard to whether one representation type arises prior to another in online linguistic production and the course of cognitive development. This sense of primary is separate from the question of whether one representation type appeared prior to the other in the emergence of symbolic communication in human history, and this claim does not necessitate or imply that one did not precede the other phylogenetically. Similarly, this implication does not follow from the fact that nonarbitrary gestural and symbolic vocal development coincide in human ontogeny. Both rely on the more general capacity to understand and use signs. Once that is in place, the ability to *acquire* either type is equipotential. However, as shown in Chapter 2, these different semiotic types can and probably do appear at different times in an *emerging* symbolic system, with nonarbitrary signs preceding and laying the foundation for conventional symbols.

To revisit the distinction between gesticulations and pantomime, it is a useful one in so far as it highlights the linguistic nature of gesticulations. However, the definition of pantomime as gesture in the absence of speech obscures the related possibility for multimodal communication involving gestures without *symbolic* vocalizations. Nothing precludes gestures that are more elaborate and pantomime-like from being co-expressed with vocalizations, particularly nonarbitrary vocalizations. Indeed, nonarbitrary vocalizations likely often do accompany gestures of this sort – for instance, imagine a child pretending to be an airplane or shoot a machine gun. The cases that McNeill (2005) focuses on appear to be exceptions and not the rule; either the vocal modality is artificially restricted (charades) or made unavailable by other atypical circumstances (a factory floor, loud concert, etc.). Given the deep anatomical integration of vocalizations and gesture (section 3.2.3), we would expect expression to remain multimodal even when it involves pantomime-like gestures. In a similar vein, although gesticulations require the accompanying words of speech in order to be fully interpretable, the same does not necessarily hold for pantomime-like gestures produced with nonarbitrary vocalizations. When shared symbolic resources are made

unavailable, the gestural modality takes on a more representationally potent quality, and there is no reason this would be impeded by accompanying nonarbitrary vocalization. Dual nonarbitrary, multimodal communication is central to the question of this thesis and will be addressed again later in this chapter (section 6).

While it may be justified to conclude that gesture is not an add-on *to* language but partly constitutive *of* language (section 3.2.4), it is not necessary to take a side in the debate over the definition of language here. We can acknowledge the semantic and anatomical relationship between speech and nonarbitrary gestures without being compelled to characterize those gestures as 'linguistic' or not. The phenomenon we are seeking to understand the origins of is the particular mode of symbolic representation prevalent in natural human communication. Having now reviewed how both modalities are typically utilized, we can accept this is accomplished primarily through vocal symbols. However, it is now clear that embodied symbolic communication cannot be limited to this single modality and sign type in isolation. The fact that symbols are intrinsically linked to nonarbitrary signs as a consequence of multimodal expression means that our explanation must encompass both of these modalities and sign types.

With these ideas in mind, we can now frame the question of where symbols come from in the context of embodied human communication. The specific question now becomes: How did the observed combined system of coordinated symbolic vocalizations and nonarbitrary gestures emerge? To begin answering this question, we must further explicate the initial conditions from which we presume emergence would have taken place, first outlined at the beginning of this thesis (Chapter 1). The following sections will seek to characterize these conditions through a brief inquiry into the evolutionary roots of multimodal communication, followed by a comparison of each modality's unique physical qualities and semiotic capacities.

4. Phylogenetic history

Previous sections have put forth evidence demonstrating that multimodality is a core feature of modern human communication. Given that the topic of this thesis reaches back through evolutionary time, it is necessary to understand the phylogenetic history

of the observed semantic and temporal coordination between modalities. As multimodality is a feature of symbolic communication, and symbolic communication is a trait that arose after the human lineage split from other primates, we must determine how it fits into the explanatory framework of symbol emergence. Put in specific terms, for the present analysis we must decide if the available evidence suggests multimodality followed symbolic communication, if it arose simultaneously, or if it preceded the development of symbols. Each scenario has different implications and raises distinct questions. If multimodality followed symbols, is it perhaps a consequence of the new mode of communicating? If the two developed simultaneously, is it an indication that the cognitive or communicative requirements of symbolic communication demand the operation of two channels of expression? If it was already in place prior to symbols, did multimodality somehow contribute to its emergence? These questions can be answered, at least in part, by looking to human evolutionary history, a major source of evidence for which is found in comparative research on primate communication.

4.1 Multimodal communication

First, we should separate two conceptually independent aspects of coordinated multimodality in language: anatomical and semantic. It is theoretically possible to have anatomical coordination – associated activation of manual and vocal systems – without semantic coordination, or even the complete absence of semantic content or communicative purpose. The simultaneous oral and manual activity observed in the early behavior of infants is an example of this latter possibility (section 3.2.3). Indeed, it seems communicative, and specifically semantic, coordination arises over the course of development from initial anatomical links (Iverson & Thelen, 1999). In addition, it may not be possible to find non-human examples of semantic coordination, as the referential status of primate communication remains in doubt (Chapter 1, section 4.3). As such, we can, for the moment, consider multimodal behavior in primates generally without the need to find semantic coordination specifically.

Before reviewing the evidence, it is worth noting that, similar to research on human communication, which only recently took a multimodal perspective, primate research

has focused largely on the vocal and manual modalities in isolation. Slocombe, Walker & Liebal (2011) argue that a long-running unimodal bias in the discipline has obscured the complexity of primate communication and the possibility that it is fundamentally multimodal. Despite the lack of attention paid to the topic, the few studies that have taken a multimodal approach provide some evidence to address this question.

In a wide review of primate research, Fitch (2006) notes that vocalizations are usually found to accompany ape manual displays. Slocombe et al. (2011) describe a number of studies on primates, as well as other mammals, that found vocal and bodily signals are produced together and integrated into complex messages. An additional and intriguing bit of evidence comes from a study on chimpanzees' atypical, voluntary vocalizations (Meguerditchian & Vauclair, 2010), which are distinguished from species-typical vocalizations that are generally innate and under less voluntary control. Meguerditchian & Vauclair observed that atypical, intentional vocalizations accompanied gestures produced while attempting to communicate with a human. In their review, Slocombe et al. (2011) conclude primate communication is inherently multimodal and note that the available data 'highlights the continuity in multimodal communication across human and primate species. This suggests that language may have evolved through an integrated combination of vocal, gestural and facial communication, rather than a unimodal system' (pg. 923).

4.2 Voluntary control and flexibility

A related issue in primate communication concerns the different levels of flexibility and voluntary control exhibited for each modality. Ape gestures are generally recognized as being culturally acquired and voluntarily produced. Ape vocalizations, in contrast, have traditionally been and continue to be viewed by many as largely innate and involuntary (a view exemplified in Tomasello, 2008), consisting of a fixed inventory of species-specific calls produced automatically in response to certain conditions. However, recent research suggests ape production may be more flexible and voluntary than once thought (Hopkins, Taglialatela & Leavens 2007; Yamaguchi & Izumi, 2008).

Regarding species-typical vocalizations, chimpanzees appear to modify their food calls according to the type and quantity of discovered food (Slocombe & Zuberbühler, 2007). In addition, chimpanzees have been shown to produce atypical vocalizations when motivated – for instance, when attempting to elicit food that is out of their reach from a human trainer (Megherditchian & Vauclair, 2010). Interestingly, exposure to a human cultural environment appears to result in innovative vocalizations. Hopkins & Savage-Rumbaugh (1991) found that, in addition to species-specific calls, the highly enculturated bonobo Kanzi produced four additional vocalizations not observed in other bonobo groups. Taglialatela, Savage-Rumbaugh & Baker (2008) later demonstrated that Kanzi's vocalizations systematically vary based on semantic context, indicating he has learned, at least to some extent, to produce vocalizations to convey referential meanings. The atypical capacities evident in these examples, if rare, nevertheless suggest intentional and flexible multimodal communication has some evolutionary precursors.

4.3 Discussion

The evidence suggests that multimodal communication has ancient phylogenetic roots – not only in the primate lineage but also apparently widespread throughout the mammalian class. The vocal and bodily modalities are used together to convey a message through simultaneous expression of related information – that is, the modalities are coordinated. The presence of this trait in nonhuman primates and other species indicates a common evolutionary origin that arose long before the human lineage split from other primates. Thus, we can assume that anatomical, and by extension communicative, coordination preceded the emergence of symbolic communication.

Although intentional and flexible vocal production appears to be diminished in comparison to that of gesture and not exhibited by all apes in all circumstances, the latent capacity for it observed in atypical cases suggests the trait, or the capacity for it, was present in the human lineage and available to be utilized if instigating conditions arose and further elaborated the behavior. Thus, we can also assume that voluntary, open-ended vocal communication accompanied the same kind of communication in the gestural modality.

These assumptions, together with the reviewed research on bodily and vocal communication in this chapter, findings from emerging sign systems (Chapter 2) and arguments for the bodily mimesis hypothesis (Chapter 1) allow us to pinpoint the situation-specific starting conditions from which our explanation of the emergence of symbolic communication can proceed. These starting conditions are the premises on which an explanation should be built. As indicated previously, we assume that referential communication is a behavior already in place. Given that the vocal and gestural modalities co-express meaning, we assume that communication is multimodal and coordinated. Given that communication is initiated through nonarbitrary signs and both modalities can be used to this end, we assume vocal and gestural signs are both nonarbitrary. In summary, the position taken in this thesis holds that the initial embodied conditions in human history out of which symbols emerged consisted of multimodal, coordinated, nonarbitrary communication.

Finally, an important implication of these premises is if multimodality is a trait that predates the emergence of symbolic communication, it is therefore not a feature that requires explanation. This interaction and synergy between modalities evident in embodied experience is an irreducible and indelible feature of communication, one that not only constrains an explanation of the origins of symbolic language, but one that can also inform it. Thus, given that multimodality was an aspect of the conditions in which symbol emergence took place, we can also consider how this factor may have contributed to the process. In other words, multimodality itself is a potential causal force and should be incorporated into our explanation.

5. Representational comparison

We have now seen how vocal and gestural forms operate in modern communication, as well as the representational qualities of each modality. The evolutionary premises outlined above entail a prior state in which symbols are not present and communication is via nonarbitrary signs expressed in both of two modalities. The endpoint of our explanation is the semiotic differentiation observed today in symbolic vocal and nonarbitrary gestural communication. A transition from nonarbitrary to

symbolic signs depends in part on how sign form features relate to meaning, and this is influenced by the modality through which signs are expressed (Chapters 1 and 2). As such, it is necessary to compare the semiotic capacities of the vocal and gestural modalities to understand how differences in the way meaning is represented could affect processes related to a transition.

First, we can note that both modalities have equal *capacity* for symbolic representation. As arbitrary signs are not bound by resemblance, they are free to take any form, which means the entire space of possible forms in both the gestural and vocal modalities is potentially arbitrary. Moreover the existence of manual sign languages illustrates that gestures can be used symbolically to the same effect as vocalizations. Thus, in theory and practice, the modalities are equivalent in the capacity to convey arbitrary signs. However, due to the distinct physical properties of gestures and vocalizations, there are critical differences in how each actualizes nonarbitrary representation and in how forms are interpreted in regard to arbitrariness and nonarbitrariness.

Our experience of the world is in large part through visual-spatial dimensions, and bodily representations are similarly visual-spatial. As a result, rich, complex and salient aspects of experience can be represented in a gesture (or combinations of gestures). Gestures can directly resemble a referent along multiple dimensions simultaneously, such as size, shape, movement, and intensity. Gestures can depict actions and events, but they can also portray static objects or entities by positioning body parts to resemble their shape or creating the outline of a form in space – and both object and action can be incorporated into a single gesture. In addition, because bodily representations have the potential to depict many different features of a given referent, the particular way one is represented can vary widely according to the demands of a particular context. This allows communicators to flexibly adjust gesture forms to highlight those features most salient and appropriate for the present situation. Finally, deixis by way of embodied pointing gestures is a powerful representational strategy, as it can identify a referent with relative precision.

Although crossmodal connections enable the vocal channel to nonarbitrarily represent features apart from the auditory, this capacity is much more limited in terms of

transparency in form (Chapter 2, section 2.2.2) in comparison to gesture. As discussed above (section 2), vocal forms can depict referents with only low precision and evoke very general qualities. Because this renders them rather blunt representational instruments, it reduces communicators' ability to craft signs that highlight salient aspects of a specific context as communicative demands shift. This lower capacity is reflected in Ahlner & Zlatev's (2010) characterization of vocal nonarbitrariness as a matter of secondary iconicity – resemblance between form and meaning is recognized only *after* the sign relationship between the two is established otherwise.

The argument that vocalization has a lower capacity for nonarbitrary representation is borne out in a recent experiment. Fay and Lim (2010) adapted the design of interactive communication studies described previously (Chapter 2, section 2.1), using vocalization and gesture as the mediums of communication. Pairs of participants attempted to communicate a list of concepts without using language in three different conditions. Participants could use nonarbitrary vocalizations only, nonarbitrary gestures only or both nonarbitrary vocalizations and gestures together. The success rate of pairs that could gesture was significantly higher than for those who could only vocalize. Interestingly, the success rate of pairs that could use both was virtually identical to those who only gestured. These results indicate that, when representation is limited to nonarbitrary signs, the gestural modality carries a markedly greater communicative load than the vocal modality.

5.1 Discussion

The first point we can note from this comparison is that, in the context of communication without symbols, gesture is a powerful and effective mode of expressing and sharing meaning. Due to the particular properties of the vocal and gestural modalities, the way in which gesture can resemble and represent referents is more varied and carries a greater amount of semantic content than that of the vocal modality. In other words, the nonarbitrary relationship between a gesture and a referent is generally more robust and direct than between a vocalization and a referent. Given that the establishment of communication in natural settings through sufficiently representationally potent nonarbitrary signs precedes and enables the

development of symbols generally (Chapter 2), we can reason that gesture, which possess these attributes to a greater degree than the other modality of communication, would have been an important factor in this process as it occurred in human evolution. Specifically, it provides additional information regarding the conditions from which symbolic communication emerged. Because gesture carries greater referential content than vocalization, we can reason that in the context of our stated initial conditions – coordinated, multimodal, nonarbitrary communication – the bulk of referential communication was accomplished through the gestural modality.

The comparison of modalities allows us to further explicate the question of the origin of symbols first refined in section 4 above. Our stated initial conditions consists of communication via a combined system of nonarbitrary vocalization and gesture, which through some means transitions to a combined system of symbolic vocalization and nonarbitrary gesture. Given that the modalities have equal potential for arbitrary representation, the question can be framed as: Why did a transition to symbols result in vocal symbols and nonarbitrary gestures instead of nonarbitrary vocalizations and gestural symbols? In theory, of course, we could ask why both are not arbitrary, or why one did not disappear as a communicative tool once symbols were established or other logical possibilities. However, psychological considerations regarding the incompatibility of simultaneous symbolic communication would preclude a scenario in which both went through a semiotic transition. Similarly, the principle of evolutionary continuity would suggest that an established – and presumably useful and perceptually salient – communicative resource (vocal or gestural) would not be lost and continue to be utilized through and after any transition in the other modality. Thus, we can justifiably limit our question to the issue of why one modality became symbolic while the other remained nonarbitrary as opposed to the opposite scenario.

These differing capacities for nonarbitrary representation and correspondingly different communicative roles have important implications regarding sign interpretation, which is a critical factor in a transition to symbolic representation (Chapter 1). The interpretive element of signs and these implications are critical to answering the question of symbols origins as it has been framed in this thesis. With the necessary premises and explanatory devices identified, the following section will now address this question directly.

6. Emergence

The ideas presented in this chapter have allowed us to frame and refine the question of the origin of symbols from an embodied perspective. We have considered the qualities of the two major modalities through which human communication takes place, the nature of symbolic communication as it exists in modern human behavior, the evolutionary history of these behaviors and the differential representational capacities of the two modalities. I have concluded that an adequate explanation of the emergence of symbols must incorporate multimodality and address how a homeosemiotic multimodal system bifurcated into the specific heterosemiotic multimodal system observed today. This section will first examine existing work on this question in light of the evidence and arguments presented here in order to establish how far we have come in formulating answers, determine if proposed solutions should be modified and identify issues that still require explanation.

6.1 Current theories

The preceding sections have highlighted the representational and communicative potency of nonarbitrary gesture and the central role it potentially played in the emergence of modern language. This idea has not gone unnoticed by other researchers, and a number of gesture-based theories of symbols origins have been proposed (Morgan, 1877; Burling, 2005; Paget, 1930; Wundt, 1916; Hewes, 1973; Corballis 2002; 2003; 2010; Gentilucci & Corballis, 2006; Arbib 2003; 2005; Arbib, Liebal & Pika, 2008; Rizolatti & Arbib, 1998; Tomasello 2008) along with, more recently, some multimodal theories (Goldin-Meadow & McNeill, 1999; Kita 2008; Kita, Kantarzis, & Imai, 2010; Kita, Ozyürek, Allen & Ishizuka, 2010; McNeill, Bertenthal, Cole & Gallagher, 2005; McNeill, Duncan, Cole, Gallagher & Bertenthal, 2008; Mithen 2004; Zlatev, 2008a, 2008b). With the exception of McNeill and colleagues, all of these have in common the premise that communication via nonarbitrary gesture preceded and enabled the emergence of symbols. Some acknowledge multimodal communication generally (Goldin-Meadow & McNeill, 1999; McNeill et al., 2005; McNeill et al., 2008; Zlatev, 2008), though only two consider the role of specifically nonarbitrary vocalization as well (Kita 2008; Kita et al., 2010a, 2010b; Mithen 2004).

These theories differ in many respects and approach the question from various perspectives. For purposes of clarity, instead of a detailed review of each in isolation, they will be addressed in aggregate in terms of the specific explanatory requirements identified in this thesis. We can assess whether they provide satisfactory explanations for the two factors underlying symbol emergence: interpretation and conventionalization (Chapter 1). Moreover, they should account for the particular version of these observed in human communication, answering i) why did vocalizations come to be used as symbols and not gesture[9], and ii) how did vocal symbols conventionalize? In addition, these explanations will need to be consilient with the premises outlined here and general evolutionary principles (see Chapter 1, section 2).

While these theories offer some insights into the question of symbol origins, a number of proposed solutions will be shown to leave critical explanatory gaps unaddressed or imply questionable evolutionary processes as solutions. For example, the questions of interpretation and conventionalization are not always addressed explicitly or as separate issues demanding their own explanations. Nevertheless, explanations concerning these are at times implicit in the arguments given and can be inferred to some extent when necessary. Finally, this discussion will not always include specific arguments from every author regarding every issue, and will instead cite a selection that is representative of general categories of similar proposals.

6.1.1 Interpretation

The question of interpretation is a question of how signs come to be perceived and used as symbols. Applying that question to human symbolic expression, we are further concerned with how vocal signs became symbolic. Existing explanations can broadly be grouped into three versions, which I will refer to as *switch*, *immediate* and *intra-modal*. Details of these versions will be addressed in the following sections, but a brief description and example of each will be useful to clarify the categorical distinctions.

---

[9] That is, the primary channel for symbols under normal circumstances, acknowledging that some gestures are arbitrary and gesture is the mode of symbolic communication in sign languages.

Switch theories assert that symbols first arose in the gestural modality and at some point the vocal modality took over the role of symbolic representation (Arbib, 2005; Arbib et al., 2008; Burling, 2005; Corballis, 2002, 2010; Donald, 1991; Tomasello, 2008). An example of a typical switch scenario is found in Arbib (2005), in which he separates the evolution of language into the following stages[10]:

1. Simple imitation
2. Complex imitation
3. Protosign
    a. Pantomime
    b. Conventional gestures (symbolization)
4. Protospeech
5. Language

Uniquely human communication first occurs through iconic gestures (Stage 3a). Symbolic communication arises when conventional gestures are invented to disambiguate pantomimic gestures (Stage 3b). Arbib also holds that communication was multimodal, and therefore vocalizations would accompany gestures (both iconic and symbolic), and these would be used to further disambiguate messages (Stage 4). Eventually, the associations built between vocalizations and gestures would allow vocalizations to convey meaning autonomously. Thus, this scenario involves the reinterpretation of gestures as symbolic as well as a subsequent interpretation of vocalizations as symbolic – that is, it entails two switches: a semiotic switch *within* one modality as well as a semiotic switch *across* modalities. The proposed switch between modalities is motivated by the fact that Arbib (2005) and other switch theory authors, for various reasons, typically exclude or downplay the role of vocalization prior to the development of symbols.

Intra-modal theories (Kita, 2008; Mithen, 2004) argue vocalizations were originally nonarbitrary and eventually became symbolic, and these theories are, not surprisingly, based on research that emphasizes the nonarbitrary potential of the vocal modality. The most well-known of this version is Mithen's (2004) proposal, which focuses on the evolution of music and language and holds that communication would have been

[10] Arbib's (2005) model also includes prior stages for grasping and mirror neurons, but these have been left out as they can be presumed for present purposes.

carried out through nonarbitrary vocalizations followed by nonarbitrary gestures. Through repeated association with communicated meanings, vocal phrases would eventually come to refer to those meanings symbolically. Thus, intra-modal accounts entail a semiotic switch within a single modality.

The *immediate* version holds that communication was multimodal and heterosemiotic at inception (McNeill et al., 2005; McNeill et al., 2008). McNeill and colleagues are the primary proponents of this model and base their ideas on work on the semiotic relationship between speech and gesture. According to this theory, there is no semiotic transition from nonarbitrary to symbolic within or across modalities, with the two modalities playing the same roles seen in modern language today.

Given the distinct but overlapping versions found in these different theoretical categories, this discussion will consider separately reasons proposed for why signs – vocal or gestural – became interpreted as symbolic, as well as reasons proposed for why symbols became the domain of vocalization.

6.1.1.1 Nonarbitrary to arbitrary

Corballis (2002) and Arbib et al. (2008) both cite the process of ontogenetic ritualization as one that could cause a transition from nonarbitrary to arbitrary signs. However, as described in Chapter 1 (section 4.3), the movements involved in this process are arguably more akin to actions, not representations. Along the same lines, any characterization of the initial movements as 'iconic' and the resulting ones as 'symbolic' is somewhat questionable and more likely an artifact of analysis on the part of human observers (as argued by Tomasello, 2008). Given the role of repeated use in altering the form of signs, and therefore potential interpretations of those signs (Chapter 2, section 2), processes similar to ontogenetic ritualization could have played some role in the development of symbols. However, because the way ritualization-derived gestures are perceived and understood is unknown, together with the fact that their representational status is questionable even for humans (Chapter 1, section 4.3), we can reason that this process does not directly speak to the critical issue of interpretation and is therefore not a satisfactory or sufficient explanation.

As mentioned above, some authors do not address the transition between nonarbitrary and symbolic signs directly or overtly in the terms of this thesis. Even so, the language used to describe a transition from nonarbitrary to symbolic signs is indicative of how it is conceived to have taken place. For example, Donald (1991) describes the development of symbols as an 'invention' to enhance conceptual abilities, a 'solution' that was conceived by an 'inventor' (pg. 219). This implies that a communicative or conceptual problem existed, and one or more individuals deliberately formulated a solution. In a similar vein, Arbib (2005) states that arbitrary symbols would have been used to disambiguate nonarbitrary representations (which are presumed to be inadequate communicative devices). Again, there is an implication that nonarbitrary communication was flawed and demanded solutions, which led individuals to develop a new communicative strategy to meet these demands. Explanations of this sort thus invoke the deliberate creation of symbolic signs by certain individuals or groups who have discerned specific advantages of communication through arbitrary signs over nonarbitrary ones. As argued in Chapter 1, however, an evolutionary explanation of the emergence of symbolic communication should avoid appealing to the insight of individuals generally, and in particular regarding the advantages of symbols, which may only become apparent after the fact. Moreover, research discussed in Chapter 2 indicates that such intentional invention does not occur even by modern, symbol-experienced populations. In addition, mimetic communication can be a remarkably effective communicative strategy (Donald, 1991; Fay & Lim, 2010), which calls into question the assumption that communicators would require or foresee reasons for any improvements. Therefore, in sum, such arguments are likewise an unsatisfactory answer to the question of novel symbol use.

Others avoid a transition in interpretation altogether and appear to assume that multimodal referential communication was a combination of arbitrary and nonarbitrary representation from inception (McNeill et al., 2005; McNeill et al., 2008). Vocalizations in these proposals are described as 'code-based' socially shared 'symbols' (McNeill et al., 2008; pg. 125). This view stems from the acknowledgment that the gestural modality has a much greater capacity for nonarbitrary representation than the vocal modality. However, in emphasizing nonarbitrariness in the gestural modality, the nonarbitrariness of vocalization is overly downplayed – both in its

potential role previous to the development of modern symbolic language as well as its remaining nonarbitrary functions today (section 2 above). Given the perceptual salience of nonarbitrary vocalizations and the fact that nonarbitrary signs are found to precede arbitrary ones across many types of emerging sign systems (Chapter 2), the assumption that vocalizations would be immediately and universally used and perceived as symbols is somewhat untenable. In related writings, McNeill and Goldin-Meadow (1999) have argued that the vocal modality was used for symbolic representation precisely because it is not amenable to nonarbitrary representation, and that symbolic function was developed to compensate for this weakness. This claim, similarly to those described above, implies insight into semiotic strategies and a deliberate effort to optimize or improve communication – in this case discerning the possibility to supplement mimetic gestures with arbitrary representations. Thus, while this view takes into consideration the distinct semiotic capacities of each modality, it dismisses the possibility and use of nonarbitrariness in vocalization and therefore does not provide a satisfactory account of how nonarbitrary signs would come to be interpreted as symbolic. Furthermore, in arguing that the vocal modality would be enlisted for symbolic representation *because of* its shortfalls, presumably to enhance communication through mimetic gestures alone, the argument is also questionable on evolutionary grounds.

Mithen (2004) takes a multimodal perspective and holds that prelinguistic vocalization was nonarbitrary, though his emphasis is on the musical and affective qualities of the modality. He claims that such vocalizations would have become associated with referents through repeated co-occurrence with meanings communicated through gesture and crossmodal connections activated by referents' features. By virtue of this association, vocalizations would eventually 'come to refer to' those meanings and 'exist as words'. Arbib (2005) also cites repeated co-occurrence as a reason for vocalizations to be imbued with meaning and become autonomous symbols. This explanation thus points to a process of ritualization that would cause initially indexical associations between non-referential forms and co-occurring referential signs (or perhaps entities physically present in the context of communication) to become perceived as symbols (akin to Deacon's (1997) claim that ritualization and indexical relationships underlie symbolic representation). One issue with this view is that it does not actually provide an answer as to why the original

perceptions of vocalizations as nonarbitrary and/or affective would be lost and reinterpreted as symbolic signs with semantic content. That is, if vocalizations were interpreted as nonarbitrary and serving some communicative function, it is not clear without additional explanation why repeated associations with certain objects or entities would lead to a change in how such vocalizations were interpreted. In addition, the characterization of nonarbitrary vocalization as non-referential[11] is not in line with the way the modality is perceived in modern use and presumably how it was so prior to symbols. Conveying affect and other non-referential information is of course an important function of the vocal modality (as well as gesture), but it is not the only one. Referential vocal signs are also an important function, one that might have been even more prominent prior to symbolic language, when the primary mode of communication was through mimetic representation. Given that nonarbitrary vocalizations are – or at least can be – *signs*, ritualization would not be necessary for them to be perceived as referential. Moreover, the expected mode of representation would have been nonarbitrary, and as argued above, ritualization is an insufficient explanation for a transition to arbitrary interpretations of previously nonarbitrary signs.

Finally, Tomasello (2008) notes the difficulties in appealing to the intentional creation of symbols and instead cites changing perceptions of signs in use as a more viable solution. He describes a scenario, previously mentioned in Chapter 2, in which individuals develop shared iconic signs (perhaps akin to the interaction-derived signs from experiments described in Chapter 2, section 2) and use them to communicate in various situations. Other individuals who happen to observe a communicative exchange involving these signs may not perceive the iconicity of the sign but could infer its appropriate meaning based on the reaction of the intended recipient. In this way, an observer could interpret the sign as representing the correct or a related meaning, but for this observer it has become an arbitrary label – or symbol. Thus, this

[11] Non-referential may be an inaccurate way to describe signs with affective content, in that they do seem to refer to certain emotional states. The more important difference between these and mimetic and symbolic signs may lie in the fact that affective signs appear to be innate, direct mappings between forms and feelings. In contrast, the form-meaning relationships for mimesis and symbols are not instinctive, arise out of experience over time and can be created and understood spontaneously based on current communicative needs.

explanation incorporates the interpretive processes underlying all sign types and the potential for sign status to transition via new or different perceptions of a sign. While these considerations are consistent with some of the ideas put forward in this thesis, Tomasello (2008) explicitly does not take a multimodal perspective as outlined here. Hence, his theory does not address the distinct semiotic capacities of each modality nor account for why certain signs over others would be perceived as arbitrary. The following section on the semiotic split between modalities will discuss these issues further. For now, we can recognize that this specific proposal offers a viable solution to the question of how symbolic interpretations of signs could arise, but not to our specific question regarding vocal symbols.

6.1.1.2 Vocal versus gestural symbols

For various reasons a number of theories postulate that symbolic communication first arose in the gestural modality, which is thought to be the modality through which the bulk of communication was accomplished (Arbib, 2005; Corballis, 2002; Donald, 1991; Tomasello, 2008). As a consequence, these *switch* versions are forced to provide an additional explanation for why the vocal modality is now the primary domain of symbols. Oftentimes purported advantages of vocal over gestural communication are cited, including lower energy requirements, the ability to communicate in the dark and freeing the hands for instruction or manufacturing. However, it is unclear whether most of these hold in practice, as sign language users are able to engage in the same activities (Emmorey, 2005) and vocal symbolic language is arguably not required for many sophisticated cultural practices – for example, building and using boats (Gil, 2006) or teaching rug manufacturing processes (Jamie Tehrani, personal communication). In addition, this solution once again relies on the insight of communicators and deliberate invention. Such assertions have also been criticized by other authors on evolutionary grounds (Emmorey, 2005), as the highly contingent nature of evolutionary trajectories means the more likely outcome of an initially gestural symbolic system would be manual signed languages, not speech. Finally, Corballis (2002) suggests the switch may have occurred because, according to him, speech is more arbitrary, and therefore the vocal modality would be more advantageous for symbolic communication. Emmorey (2005) counters this line of reasoning, noting the many arbitrary signs found in signed languages and the absence of any evidence demonstrating that expression or

processing is hindered by nonarbitrary manual signs. On general semiotic grounds, this suggestion also does not hold. As discussed in section 5 above, vocal and gestural signs have an equal capacity for arbitrary representation, which means neither has a theoretical 'advantage' over the other for conveying symbolic signs. While this suggestion acknowledges the modalities' distinct semiotic qualities, it does not go on to consider how interpretive processes would have affected the way signs in each modality were perceived and used, and therefore does not constitute a sufficient explanation.

Some *intra-modal* theories do not provide a reason for why one modality is now predominantly symbolic and not the other (Kita, 2008; Kita et al., 2010a, 2010b; Mithen, 2004). Zlatev (2008) briefly mentions Corballis's (2002) notion that the greater arbitrariness of the vocal modality as a possible cause, though the idea is not developed further. Thus, while these theories are consistent with our premise of multimodal communication and avoid criticisms related to a switch, they do not directly address the question of heterosemiotic distribution across the modalities in regard to sign interpretation.

The *immediate* theory (McNeill et al., 2005; McNeill et al., 2008), which assumes vocal signs were *de facto* symbols, seems to suggest that vocalizations were used as symbols *because* they are arbitrary. In other words, because the modality could not be utilized effectively for nonarbitrary representation, vocal symbols were created in order to make the modality communicatively useful and match the referential capacity of gesture. Again, this account appeals to the ability and inclination of communicators to identify potential improvements and design novel strategies to optimize communication. Moreover, it does not answer *why* vocalizations were interpreted as arbitrary (as opposed to gestures) and simply assumes that they would be.

6.1.1.3 Summary

This review and discussion has revealed that the question of interpretation has not been thoroughly or sufficiently addressed in existing theories of symbol origins. The explanations provided prove to be incomplete or otherwise unsatisfactory for a number of reasons, including by adopting premises not in line with those identified as

most appropriate here, not accounting for all the critical factors underlying symbolic communication identified in this thesis, invoking intentional acts that lack viability and/or misapplying or failing to apply concepts from semiotics. Tomasello (2008) provides the groundwork for a solution by pointing to the potential for intended nonarbitrary signs to be mis- and reinterpreted as symbolic by certain individuals – a viable possibility made evident in the research discussed in section 4 above. However, relevant issues related to sign interpretation are left unaddressed, and as a result, a number of questions remain concerning a transition to the particular heterosemiotic, multimodal communication observed today.

With the contributions and deficiencies in existing theories identified, a more comprehensive explanation can now be formulated that incorporates and makes use of the actions of interpretive processes and the consequences of each modality's particular semiotic properties relative to each other. After a similar review concerning the question of conventionalization, this will be presented in section 6.2 below.

6.1.2 Conventionalization

The question of conventionalization has received even less explicit attention and analysis than interpretation. It may be that some authors assume the issue has already been resolved by computational models that demonstrate populations of agents can converge on shared vocabularies through communicative processes (Chapter 2, section 4). However, as discussed in Chapter 2, the success of these models requires that specific conditions hold – namely, some reliable source of information must provide communicators with access to each other's intended meanings, thereby grounding new arbitrary signs. While cross-situational learning has been shown to aid this process to some extent when a meaning intended to be conveyed is not entirely certain (A.D.M. Smith, 2001), its effects are weakened as referential uncertainty increases (Smith et al., 2011). In addition, cross-situational learning is perhaps better characterized as an extra-communicative process that can support the acquisition of a vocabulary as opposed to a direct sign-grounding mechanism in itself. In settings like those of emergence, uncertainty-lessening mechanisms internal to communicative acts like establishing joint attention are of central importance for ensuring symbols are grounded and conventionalization takes place. Thus,

conventionalization should not be considered an automatically guaranteed outcome, and a satisfactory explanation for symbol origins in human evolution as offered in these theories should cite some mechanism or mechanisms that would accomplish this grounding.

At times, some authors seem to conflate conventionalization with arbitrariness. For example, likening pre-linguistic gestures to those of primates, Arbib and colleagues (Arbib et al., 2008) claim that ritualization processes would have led initially iconic gestures to be conventionalized and symbolic. Similarly, Corballis (2002) argues that iconic gestures would become progressively abstract and symbolic while becoming conventionalized. However, conventional signs are not necessarily symbolic, and the alterations in sign form resulting from ritualization and interaction processes may not correspond to a change in sign interpretation (Chapter 2, section 2). Moreover, conventionalization through ontogenetic ritualization occurs between an isolated pair of individuals and does not typically spread to others. Therefore, this process would probably not play a major role in the conventionalization of arbitrary signs throughout a population. Given that such processes do not account for transmission to others or a grounding mechanism underlying the large-scale spread of novel arbitrary signs, we can conclude this explanation does not sufficiently address the question of conventionalization.

Many authors, as with interpretation, address the issue of conventionalization indirectly or only implicitly. Kita and colleagues (Kita 2008; Kita et al., 2010a, 2010b) argue that crossmodal nonarbitrariness would allow communicators to agree upon vocal labels. The fact that such labels might not actually be symbolic in the first place aside, this account again relies on the intentional acts of communicators – in this case, the deliberate construction of shared vocabularies. Or, if what is intended in this line of reasoning is that shared crossmodal nonarbitrariness would lead to communicators sharing vocal labels *without* the need for explicit agreement, the issue arises that conventionalization processes for arbitrary signs may differ from that of nonarbitrary signs. In which case, convergence on shared signs due to common sensory associations is probably not an appropriate model and likely cannot account for convergence on shared symbols. In another vein, McNeill and colleagues (2008) seem to imply that communicators would explicitly negotiate and agree on labels, as

no other explanation is given for how a code-based vocal system would come to be established. For the same reasons cited previously, however, appeals to such intentional mechanisms are not a viable solution.

Finally, Tomasello (2008) argues that conventional signs would develop from interlocutors' imitation of each other's iconic signs, which could then spread to other individuals who observe communication via these signs. While this account provides a solution to how a sign could spread beyond a dyad or small group, some issues arise when we consider how this process might explain population-wide conventionalization. The grounding mechanism that enables a hypothetical observer's initial reinterpretation is indirectly provided by another's correct recognition of an intended nonarbitrary gesture and their subsequent actions. If we extend beyond these highly constrained circumstances, however, such information would not be available as a grounding mechanism if the observer attempted to communicate with a new individual using the gesture. Neither information from nonarbitrary representations nor inference from others' successful communication via nonarbitrary representations would be present to aid interpretation. As described in Chapter 2, section 2.2.1, inference from ambient sources on its own can assist this process to some degree. However, it is a relatively weak mechanism compared to other grounding sources and we should not assume such factors would be consistently available or sufficiently informative. Therefore, implicit reliance on contextual inference as the primary means of grounding novel symbols in this way does not provide an adequate explanation for conventionalization.

A related scenario that might be put forward from this view is the possibility that it would occur with a *group* of observers, perhaps children. These observers could then use the gesture between themselves to communicate this meaning, and their behavior could then spread to others in a population who observe their interactions. The problem with this scenario is that it requires both i) that multiple observers fail to recognize a gesture as iconic and ii) that all those who have done so assign the same or very similar meanings to the now arbitrary gesture. Otherwise, the gesture would not be a functional communicative device between these individuals, despite their shared experience. As this outcome requires that the same set of complex psychological processes occurs simultaneously across multiple individuals, it is a

highly unlikely possibility. Thus, while Tomasello (2008) addresses how a novel symbol is *established* (Chapter 2, section 2.3) for an individual, without any other means of providing reliable information to indicate meaning for others, this explanation does not fully account for how arbitrary signs could be *grounded* for others and conventionalize throughout a population.

6.1.2.1 Summary

The question of conventionalization demands an explanation of what factors could have grounded novel symbols. Existing theories either provide no account at all, point to deliberate construction, or rely heavily on weak forces. As with interpretation, then, substantial issues remain as to how conventionalization might have taken place.

6.1.3 Discussion

Together, existing origins theories go a long way in explaining how human communication evolved. As a group, they recognize the necessity for a prior nonarbitrary mode of communication in establishing a symbolic system. Multimodal theories address language as a coupled vocal-plus-gesture communicative system. Nevertheless, existing theories do not provide a complete explanation for the emergence of symbolic communication according to the criteria put forth here as most critical. These theories thus far either have not addressed the question of symbol origins in these terms, have done so only superficially and/or invoke mechanisms of questionable validity to explain the creation and/or conventionalization of novel symbols. Furthermore, processes of sign interpretation and consequences of the particular qualities of embodied human communication have not been fully mined to exploit their inherent explanatory potential.

With the questions themselves, our explanatory framework and the current state of theoretical progress now comprehensively established, the following section will address symbol origins in human evolution with these ideas together in mind.

6.2 A model of emergence

The question of symbol origins in the context of interest here hinges on i) how nonarbitrary signs became arbitrary and ii) how arbitrary signs conventionalized. I have argued that answering these questions demands consideration of semiotic concepts, evolutionary principles (Chapter 1), general processes in emerging sign systems (Chapter 2) *together with* the particular features of embodied human communication (sections 2-5 above). I will propose explanations for these two factors using the evidence presented and the concepts and principles developed thus far in this thesis, arguing that this perspective avoids problematic assumptions and implications and fills explanatory gaps in existing proposals.

6.2.1 Interpretation

As described in Chapter 1, symbolic interpretations of signs can arise in modern human communicative processes simply through a failure to recognize an intended nonarbitrary relationship between form and meaning. In other words, unintentional *mis*perception can lead to symbols without the need for individual insight or deliberate creation. As argued in section 3.2.7 above, we can assume communicators engaging in bodily mimesis in the context immediately preceding symbols would possess a cognitive capacity for referential representation encompassing all sign types. Therefore, this potential for symbols as misperceived nonarbitrary signs existed in our hypothetical circumstances. Given the unviability of mechanisms based on deliberate invention, before reverting to appeals to them we should exploit the explanatory power of this element of signs to its full potential. Thus, without evidence demanding the need for additional justifications, I am in agreement with Tomasello (2008) that a transition from nonarbitrary to symbolic was a consequence of misinterpretation.

Apart from a general explanation for symbols, however, Tomasello's account does not offer a satisfactory explanation for why *vocalizations* are now the primary domain of symbols. This is due to his assumption that only gesture would be used for intentional, referential communication, which ultimately negates – or at the very least vastly diminishes – the possibility to explain why one modality came to dominate symbols without appealing to less plausible evolutionary and psychological processes.

That is, if we presume arbitrary signs arose in gesture because it was responsible for referential communication, we must then justify i) why the vocal modality would become voluntary and flexible and ii) why symbols would dominate it and not the modality in which they were already utilized. As argued above (section 5), however, there is strong evidence to suggest referential communication at this point in human history was multimodal. I will argue that taking this multimodal perspective together with explicit acknowledgment of sign interpretation processes can go beyond an explanation for the emergence of symbolic signs generally and inform the related and more specific explanation for the emergence of vocal symbols.

Although the interpretive nature of signs provides an opportunity for a semiotic transition, that transition will depend on the specific semiotic and communicative characteristics of the representations and the channels through which expression takes place. As described above, the modalities through which humans communicate have distinct semiotic profiles – namely, gesture has a greater capacity for nonarbitrary representation than vocalization, and as a result, gesture presumably would have been largely responsible for accomplishing referential communication. These differentiated representational capacities and communicative functions entail that the nonarbitrary relationship between a gesture and its intended meaning is more likely to be perceived as such than that of a vocalization. In other words, vocalizations are more apt than gestures to be *mis*interpreted as arbitrary. Moreover, we can reason that misperception of a vocalization would not substantially disrupt communication, as the bulk of an intended message would be conveyed in and understood through gesture. Thus, sign reinterpretation and the establishment of symbolic reference could arise through natural processes inherent to any communicative exchange. Acknowledging these aspects of embodied multimodal communication makes available a solution in which vocalizations take on a new semiotic function without invoking the deliberate actions of communicators to discern the potential for and create symbols.

Importantly, this argument does not imply that *only* vocalizations would have been misinterpreted as symbols. It is of course possible for nonarbitrary gestures not to be perceived as intended, and this no doubt took place frequently. The crucial factor is the *relative* occurrence of this between modalities. Lower nonarbitrary potential would lead more reinterpretations of vocal signs to arise than gesture. We would

expect this larger proportion of individual, vocal symbolic interpretations to be amplified through cultural processes and eventually result in a greater proportion of symbolic vocal signs across a population. This outcome is precisely what is observed in language today, with the vocal modality primarily – but not solely – symbolic, and gesture primarily – but not solely – nonarbitrary.

It is also important to note that a sign being reinterpreted as symbolic for one individual does not necessarily mean others will also perceive it as an arbitrary symbol when that individual later uses it in communication. Nevertheless, at least two factors suggest this would occur with some regularity in the case of vocalizations. First, the fact that communicators share perceptual biases increases the likelihood that what one individual interprets as arbitrary will be similarly interpreted by others – other factors like interactive history and common ground being equal. Second, the low nonarbitrary potential of vocal forms would minimize any interference that might arise from possible nonarbitrary interpretations for a given context of use. That is, this low potential means that if a vocalization could potentially be perceived as nonarbitrary to a particular referent, that potential connection would be relatively weak and therefore not perceived readily. These factors of course do not guarantee novel vocal symbols would automatically be interpreted as such, but their overall effect would be to encourage similar perceptions.

Aspects of the ideas outlined here may appear at first to be very similar to McNeill & colleagues' proposals (section 6.1.1). Their position holds that the particular heterosemiotic speech-gesture system observed today is in part due to the modalities' distinct semiotic profiles. However, that view differs to the one presented here in both i) underestimating the potential for and perceptual salience of nonarbitrariness in the vocal modality, and ii) implicitly adopting the questionable assumption that communicators would seek to optimize the communicative potential of each modality and be able to discern how to do so. Regarding i), the vocal modality has nonarbitrary potential beyond direct iconicity by virtue of shared crossmodal associations, and nonarbitrary vocal signs are both incorporated into language and created on the fly when the need or inclination arises (section 2 above). The capacity and pervasive use of nonarbitrary vocalization in modern behavior together with research indicating communication is established through nonarbitrary representation in natural settings

(Chapter 2) is reason to expect the vocal modality to, like gesture, first be used for nonarbitrary signs. This in combination with factor ii) serves to obscure the possible role of unintentional interpretive processes in rendering vocalizations symbolic and may explain the assertion that the modality would have been used symbolically from the earliest stages of referential multimodal communication. The reasoning appears to be that, because the vocal modality was so ineffective at nonarbitrary representation, it was put to use by communicators in a way that did not require that capacity and also complemented gesture. It may be that the authors do not intend to make these specific claims as made explicit here, but their language and arguments are at the very least suggestive of them. I contend that explicitly citing unintentional interpretive processes and accepting a more accurate characterization of multimodal communication that recognizes vocal nonarbitrariness clarifies the issue and avoids any potential misunderstanding or flawed implications that could result otherwise. On the present view, then, the limited nonarbitrary capacity of vocalizations spurs greater *mis*perceptions of symbols relative to gesture, which ultimately translates into the heterosemiotic system observed today. In this way, vocalizations are rendered symbolic by the operation of blind psychological and large-scale cultural processes over time.

In a related vein, Corballis' (2002) suggestion that vocalizations would be used for symbols because of the modality's more arbitrary nature is redolent of similar issues. The emphasis is placed on arbitrary capacity, the implication being that this would make vocalization more advantageous for conveying symbols. However, as discussed above, the two modalities are equal in regard to capacity for arbitrariness – where they differ is in nonarbitrary representation. With this difference alone, the semiotic transformative potential of interpretive processes is sufficient to explain why vocalizations became predominantly symbolic and not gestures.

I have argued that incorporating sign interpretation and embodied human communication into our explanatory framework reconciles problems in existing symbol origins theories. Doing so also allows us to formulate a more comprehensive and theoretically satisfactory account of the emergence of arbitrary vocal signs that maintains evolutionary and psychological plausibility. These ideas, their implications

and related questions will be discussed further following a similar treatment of conventionalization.

6.2.2 Conventionalization

A shift in a population-wide communicative strategy from nonarbitrary to symbolic representation poses a substantial problem (Deacon, 1997; Donald, 1991). Novel symbols resulting from individual reinterpretations will not be shared by others and will therefore be a less effective communicative resource than nonarbitrary signs, which are readily interpretable without knowledge of arbitrary form-meaning mappings. Given this inherent reduction in communicative efficacy resulting from the initial use of novel symbols, and no previously established symbolic system or an expectation to acquire one, how would individually held arbitrary signs be understood and adopted by others? Once nonarbitrary information contained in signs is no longer a reliable indictor of meaning, some other source must compensate for this loss in order for novel symbols to be interpreted as intended and adopted by others. Thus, an account of the emergence of conventional symbols should explain how the tension between using novel symbols and successful interpretation of signs is resolved.

The theories reviewed above were found largely to disregard the question of conventionalization or offer unsatisfactory accounts. Tomasello (2008) describes how a sign form used between certain interlocutors could be adopted by an observer, and such a scenario likely would have been frequent and responsible for the spread of sign forms as symbols to some individuals. However, as argued above, it does not adequately explain how a newly interpreted symbol might move beyond these highly constrained circumstances and propagate through a population.

In the same way that the origin of symbolic interpretations is informed by the ideas and evidence outlined in this thesis, the question of conventionalization similarly benefits from and demands consideration of these issues. In addition to embodied, multimodal communication, we can also incorporate evidence from emerging sign systems, the most relevant being computational models (Chapter 2, section 4). While models have demonstrated shared vocabularies can arise through processes of exchange and learning, this only occurs when certain conditions are met.

The critical requirement for conventionalization is that symbols are grounded – that is, recipients are able to identify an intended referent and thereby adopt the correct form-meaning mapping (Chapter 2, section 4.3). Modelers point to joint attention (Steels, 2003; Steels & Kaplan, 1998) together with cross-situational learning strategies (A.D.M. Smith, 2001; Smith et al., 2011) as mechanisms that would enable grounding and shared symbols in real-world settings. While some models' success has rested on simulating perfect joint attention (Steel & Kaplan, 1998; K. Smith, 2002), those modeling cross-situational learning indicate attaining successful convergence can tolerate some level of referential uncertainty (A.D.M Smith, 2001). Nevertheless, weaker cross-situational learning strategies are employed in environments like those presumed for emergence (Smith et al., 2011), and its power is also diminished as uncertainty increases (Smith et al., 2011; Vogt & Couman, 2003). In consequence, establishing and maintaining joint attention remains integral to the conventionalization process. Therefore, an adequate explanation demands a mechanism or mechanisms to achieve this.

Again, we should avoid citing overt instruction or agreement as cause for shared arbitrary mappings to arise. As mentioned in Chapter 2, a number of other socio-pragmatic factors that do not rely on deliberate action are thought to underpin joint attention and symbol grounding in language acquisition (Tomasello & Farrar, 1986) including gesture recognition, eye-gaze tracking, shared task awareness, script execution and recognition, and emotion recognition and synthesis (Steels, 2003). We can therefore look to how these more plausible mechanisms might have operated in symbol emergence and if they could have enabled conventionalization to take place.

We can reason that most if not all of the joint attention aids listed above would be present in our hypothetical context, given the requirements and consequences of intentional, referential communication through mimesis. Notably, gestural and bodily communication feature prominently in this list, and some researchers specifically cite pointing gestures as their presumed grounding mechanism, which is implemented as perfect context reduction (Steels, 2003; Steels & Kaplan, 1998; Chapter 2, section 4). While pointing is a particularly effective and direct type of nonarbitrary gesture, the preceding discussion of human multimodal communication shows that we should also consider the potential role of iconic and icon-index hybrid gestures as information-

rich sources that could indicate communicators' intended messages, particularly as identifying a referent through pointing will not always be an available strategy when communication is about referents displaced in time and space from the current context.

Multimodal communication as described here involves the simultaneous expression of representations in two channels. In the conditions preceding conventionalization, the utterances of interest presently are those conveying novel vocal symbols. We would expect these vocalizations to be produced together with some sort of gestural sign. Furthermore, we can assume such gestures would frequently be manifested as nonarbitrary representations depicting a meaning related to the intended meaning of the accompanying vocal symbols. In addition, given the representational capacity and perceptual salience of nonarbitrary gestures, we can reason receivers could reliably recognize gestures as intended – at least in part – and use them to interpret multimodal messages. Likewise, we can posit that an intended arbitrary vocalization would not interfere with comprehension enough to impede communication; because of its relatively low representational capacity, any potential nonarbitrary incongruency between a vocalization and referent would be eclipsed by the referentially informative representation conveyed via gesture (indeed, a perceived mismatch between modalities may itself ensure novel vocal symbols are correctly interpreted as arbitrary). In this way, the coordination of vocalization and gesture makes it possible for novel symbols to be used while maintaining effective communication. The information made available in referentially potent gestures compensates for the presumed lack of information contained in the vocal modality, which allows others to recover intended messages and potentially adopt symbolic interpretations of vocalizations. Thus, multimodal representation of the type embodied in human vocal and gestural communication provides a stable, organic grounding mechanism that could support the conventionalization of vocal symbols.

As mentioned above, the initial use of arbitrary signs not yet shared by others is a less effective communicative strategy compared to nonarbitrary representation. The inherent tension that arises in a transition from communication through nonarbitrary signs to conventional symbols can be resolved by considering how information is conveyed simultaneously by the two modalities. The gestural modality is

disproportionately responsible for nonarbitrary referential communication. This asymmetry both fosters arbitrary interpretations of vocalizations relative to gesture and ensures their use will not compromise communication. As a result, the shift to what would initially be an inferior communicative strategy in one modality is enabled by the presence and qualities of the other modality.

A useful analogy to draw is to the concept of genetic shielding (Deacon, 2009). Genetic shielding occurs when multiple causal sources initially serve the same function – for example, an organism that possesses both a gene responsible for manufacturing a certain nutrient and a regular food source that supplies the same nutrient. Given this redundancy, it is possible for the gene to mutate because the presence of the food behavior ensures the original function continues to be served. The mutated gene may become nonfunctional or it may take on a new useful function in some other capacity. Thus, by one source remaining devoted to the genes' original purpose, an opportunity arises for that gene to develop a new one. In nonarbitrary multimodal communication meaning is expressed through two channels at the same time to convey an overall message, creating a situation similar to genetic redundancy. If arbitrary vocalizations began to be used, nonarbitrary gestures would continue to be produced and available to aid interpretation. Given gesture's high representational capacity, it would be capable of supporting communication, to a degree, despite the loss of nonarbitrary information in the vocal modality, which would in turn enable the spread of accompanying vocal symbols to others. Like the shift made possible through genetic shielding, one modality is free to undergo a semiotic transition because the other modality continues to carry out the function of communication while the new form of representation is adopted by others and becomes an effective communicative device in its own right.

This analogy is of course not to dismiss signed languages, wherein conventional symbols arose in one modality alone. Interestingly, however, signed languages are marked by having a large proportion of nonarbitrary signs (Perniss et al., 2010), while speech is characterized by a higher proportion of arbitrary words. In addition, sign languages exhibit a lower degree of conventionalization than spoken language (Chapter 2, section 3). In other words, typical symbolic communication as observed in language is unlike signed language in being a highly conventionalized system of

predominantly arbitrary signs. These differences may result from sign languages being unimodal and the particular features of the modality in which it takes place. First, we can note that signed communication is through a modality with a high capacity for nonarbitrary representation. It may be that fewer signs are arbitrary in part because nonarbitrariness in gesture is easily recognized. If fewer signs are misperceived overall, then we can reason arbitrary gestures would not come to dominate the modality as arbitrary words do the vocal modality. Another factor may be that the use of novel arbitrary signs has a greater negative impact on unimodal communication, as it automatically knocks out the only channel through which meaning would normally be shared through nonarbitrary signs. Of course, the capacity and effectiveness of gesture can still provide a larger communicative 'superstructure' that supports joint attention and thereby conventionalization potentially to some degree, which could explain the emergence of a more semiotically mixed system like that seen in sign languages. Nevertheless, it may that the greater potential for novel arbitrary gestures to disrupt unimodal communication than novel arbitrary vocalizations to disrupt multimodal communication can account for the observed differences with vocal language. Thus, while conventional symbols can arise in a single modality, the presence of two channels with unequal representational capacities could be responsible for the more thoroughly arbitrary and conventional system found in speech.

A point to emphasize is that this account holds that conventionalization of vocal symbols could occur through cultural processes like those modeled in computational simulations due to factors inherent to communication as it naturally takes place. By recognizing that human communication is multimodal we are able to explore the consequences and consider its potential explanatory contribution. Specifically, the presence and coordination of two information channels provides a built-in grounding mechanism, making it possible for conventional symbols to emerge without intentional construction.

The ideas outlined here offer a more explicit and comprehensive explanation for conventionalization than existing theories. Importantly, this proposal does not rely on insight and factors related to deliberate construction. The next section will discuss implications and further questions that come out of this proposal.

6.2.3 Discussion

I have argued that novel symbols would have arisen, not through creation in response to a discerned need or pressures for improved communication, but unintentionally when nonarbitrary signs are misunderstood – an inevitable consequence of any imperfect communication system. Furthermore, I have argued that those symbols would have been vocal neither because the modality is a superior symbolic medium nor as a result of communicators' awareness of its inadequacy as a nonarbitrary medium and their intentional attempts to develop a more effective communicative strategy. Instead, the vocal modality would have become predominantly symbolic because its lower nonarbitrary capacity increases the likelihood that vocalizations are perceived as arbitrary. In addition, I have argued that vocal symbols could have become shared and conventionalized because multimodal communication provides a grounding mechanism 'for free', which allows populations to coordinate arbitrary form-meaning mappings through cultural processes without overt agreement or construction. The coordinated vocal speech and nonarbitrary gesture mode of communication observed today is, on this account, simply an epiphenomenal consequence of imperfect communication as it is carried out in embodied human experience.

While indiosyncratic symbolic interpretations of signs will inevitably result from imperfect nonarbitrary communication, this does not necessarily entail the establishment of conventional symbols. If conventionalization processes were not sufficiently supported by symbol grounding sources, shared symbol systems conceivably might not emerge, with communication remaining predominantly nonarbitrary. In the case of symbols in human evolution, the factors underlying conventionalization are put in place and strengthened by virtue of the particular characteristics of human communication. The way that *mis*interpretation of nonarbitrary vocalizations can become *re*interpreted symbols that subsequently become shared across a population as described follows directly from the low nonarbitrary potential of vocal forms, the high nonarbitrary potential of gestures and their simultaneous expression. In this sense, the potential for vocal symbolic communication to emerge through cultural processes stems from and is supported by the distinct embodied conditions of human communication, not from inherent superiority or a drive for progressively more complex systems of communication.

A point worth noting is that the transition from nonarbitrary signs to symbols in the context of multiple channels of expression as described here differs from the way it typically has been conceived thus far. In research on emerging sign systems (Chapter 2, sections 2 and 3) and many symbol origins theories the focus has been on how nonarbitrary signs could ground communication, and then how *those same signs* could become symbolic. For example, in graphical communication studies, experiments involve looking at how drawings are created and change over time through use. Similarly, a number of origins theories involve nonarbitrary gestures bootstrapping the development of symbolic gestures (see theories discussed in section 5 above). When the process is framed in terms of multimodal communication, however, the way a transition might play out changes, as either or neither channel could in theory become symbolic. In consequence, it becomes crucial to consider the semiotic properties and communicative contributions of the modalities relative to each other. In the conditions of human multimodal communication, gesture would have been primarily responsible for establishing communication, but, as argued here, it would be the vocal modality that underwent a transition to predominantly symbolic communication as a result of how vocal signs would be perceived relative to gesture. In this case, signs that enable effective and robust communication through nonarbitrary representation that could underpin the development of symbols can continue to serve that function even as symbols develop; because a separate channel is available that is not as vital to referential communication, symbols can emerge in it without sacrificing the powerful communicative contribution of nonarbitrary representation in the other channel. Thus, multimodal communication creates opportunities not available in unimodal systems. Exploring that potential as I have here makes it possible to account for the emergence of symbolic communication without extraneous appeals to external factors.

One final point worth addressing is the changing role of gesture during and/or after the development of a fully symbolic vocal language. Co-speech gestures (*gesticulations*, in McNeill's (2005) terms) are typically less semantically specific and elaborate than pantomime-like gestures, which would have presumably been in greater use in our hypothetical pre-symbol environment. The different characteristics of these two gesture types may arise simply from shifting cognitive demands that

would have resulted from the vocal modality coming to dominate referential communication after an extensive symbolic language was constructed. If greater attention and cognitive resources would need to be devoted to accessing and using a learned, symbolic vocal code, the level of overt attention devoted to the gestural modality may decrease in response. Based on the principle of continuity (Chapter 1, section 2) we would expect communication to remain multimodal and gestures continue to serve a similar function, and the shifting communicative weights and cognitive demands that would occur after the development of symbolic language could account for the nature of gesture as it is produced in accompaniment to speech.

Donald (1991), Zlatev and colleagues (Zlatev, 2008; Zlatev et al. 2005), and Tomasello (2008) have argued that the abilities and behaviors underlying bodily mimesis are a necessary step in bridging the gap between primate communication and human language. Given what is known regarding the biological and cultural differences between primates and humans, we can conclude that the evolution of bodily mimesis in the human lineage would have involved biological evolutionary forces as well as biological and cultural co-evolution. The proposal here for bridging the gap between bodily mimesis and symbolic communication, however, suggests that we may not need to posit additional or independent causal forces to justify the emergence of symbols. In other words, once the capacity for intentional, referential nonarbitrary communication through vocal and gestural modalities was in place, the development of conventional symbolic communication is achievable through 'blind' cultural processes alone. This position adheres more closely to the principles of parsimony and continuity than existing symbol origins theories, but it also has important implications for the larger study of human cognitive evolution. If, from the premises set out in section 4 above, coordinated symbolic vocal and nonarbitrary gestural communication is a potential direct consequence of psychological and cultural processes intrinsic to communication itself, it suggests we are not compelled to propose specific events or circumstances that would have created pressures for symbols. In this case, the power of multimodal bodily mimesis and related capacities is revealed to have great evolutionary significance and important behavioral and cultural consequences, one of which, as I have argued, is the potential to establish conventional symbols. With the need for additional explanations negated, focus can be placed on the foundational abilities and behaviors that create the conditions in

which emergence can take place (Zlatev et al. 2005 discuss the evolutionary transition from primate cognition and bodily mimesis and provide a starting point for this line of research). Thus, taking this perspective allows us to identify factors that represent significant turning points in human evolution and thereby focus our research efforts to reach a better understanding of how evolutionary forces could produce the human mind and culture.

These arguments are not intended to suggest that all features of symbolic modern language can be accounted for by the conditions and processes outlined here. Specialized phenomena like phonology and syntax involve additional factors, which are beyond the scope of the present discussion. A large body of research has taken conventional symbolic communication as a basis and addressed how particular aspects of language could arise, including compositionality (Kirby, 1999; Kirby et al., 2008) and phonological structure (de Boer, 2000). The present proposal locates this basis in evolutionary terms and provides a foundation for this type of research.

Finally, a number of empirical questions are raised by the specific arguments and claims presented here. I will address two immediately apparent questions, one concerning interpretation and the other conventionalization.

In regard to interpretation, I have argued that vocal signs that were not recognized as nonarbitrary would, at least on some occasions, be perceived as symbolic and adopted as such even without the expectation to learn an established system. This implies that, upon first hearing, what were perceived as arbitrary vocal forms would be attended to, understood as meaningful and remembered. This is based on evidence from gesture and language research that indicates people integrate signs from the two modalities into a psychologically unified utterance, which suggests vocal and gestural signs would be processed with similar parity and coordination in pre-linguistic multimodal communication. It is also possible, however, that the way each modality was perceived would be altered by the asymmetry in communicative content characteristic of the circumstances preceding conventional speech. That is, the saliency, referential potency and resulting communicative dominance of gesture could itself impede the emergence of vocal symbols through unintentional cultural processes alone. If gestures were primarily responsible for conveying messages, then what would be

largely uninformative perceived arbitrary vocalizations could be disregarded with little effect. Indeed, the focus on gesture might distract recipients from vocal signs and hinder interpreting and learning them as symbols in their own right. If this were the case, it would suggest the reasons for the emergence of vocal symbols put forward here are insufficient and we may in fact need to reconsider other factors in human history, including those asserted in existing theories, that could have caused vocalizations to be used as symbols. The following Chapter will be devoted to an experiment designed to test the predictions that follow from these competing positions.

In the case of conventionalization, I have argued that nonarbitrary gestures would have provided sufficient grounding for arbitrary vocalizations. This implies that the kind of information made available in nonarbitrary gesture could indicate meaning with enough reliability to underpin convergence on shared signs in a population. Computational models have demonstrated that convergence can be reached when agents have unambiguous knowledge of intended referents, but such perfect circumstances lack ecological plausibility. Other studies show that in the presence of some referential uncertainty convergence can be aided by cross-situational learning strategies, but mechanisms to constrain context sizes are necessary to ensure conventionalization is robust and reliable. Information provided by gesture is only partial and does not indicate referents with perfect accuracy. It could, however, serve to reduce the number of potential referents sufficiently to support conventionalization, though this possibility has not yet been tested. In addition, models have not implemented multimodal signaling, nor, more specifically, the possibility for two types of representation and information. It remains an open question whether partial information like that embodied in gesture and the presence of two signal channels could support convergence. I will present a simulation designed to address these questions in Chapter 5.

7. Conclusion

Human communication is carried out in two modalities (primarily) utilizing both nonarbitrary and arbitrary representations, though the distribution of semiotic types is uneven across these modalities. Specifically, it is characterized by predominantly

symbolic vocal speech coordinated with predominantly nonarbitrary gesture. It is this heterosemiotic, multimodal nature of modern language that is both the phenomenon to be explained as well as an explanatory resource in itself. Existing theories that seek to explain the emergence of symbolic communication have either not provided a comprehensive account or implied – whether intentionally or not – the action of mechanisms and forces that are inconsistent with the foundational ideas and principles outlined in Chapter 1. A more complete, theoretically satisfactory and clear explanation can be formulated by adopting an embodied perspective and using the broader range of available evidence reviewed in this Chapter (sections 2-4) and Chapter 2. I have argued that symbolic vocal communication coordinated with nonarbitrary gestural communication is a highly probable trajectory for what would have begun as a homeosemiotic, multimodal communication system. This outcome could arise through unintentional psychological and cultural processes, and thus the explanation provided here avoids problems associated with previous proposals.

CHAPTER 4
Interpretation: Experiments

1. Introduction

This chapter will investigate one of the specific claims regarding sign interpretation raised in the previous chapter. First, I will review research that can inform our understanding of how vocal and gestural signs are processed and provide an experimental framework for investigating the questions at hand. I will identify specific hypotheses to be tested to address these questions and describe a series of experiments designed to do so. Findings will be discussed in regard to the focus of this thesis as well as their implications for the larger study of language and gesture, including avenues for future research.

2. Background

A number of recent studies have examined the relationship between speech and gesture. These typically involve a comparison of how messages conveyed through unimodal or multimodal utterances affects how that information is interpreted and processed.

Kelly et al. (1999) showed participants videos of speakers conveying multimodal utterances that included non-redundant information in gesture. When asked to recall the speaker's words specifically, participants referred to meanings expressed only in gesture, though many did not recall getting any information from gestures, indicating signs from the two modalities are integrated unconsciously and automatically. In a follow-up study, Kelly et al. (2010) again found that gesture affects speech comprehension in this way but also that speech in turn affects gesture comprehension, demonstrating that inter-modality influence is bidirectional. In addition, they found that gesture's effect on speech comprehension persisted even when participants were told to attend to verbal information only, indicating that speech-gesture integration is an automatic, subconscious process.

In a similar vein, McNeill et al. (1994) investigated how listeners interpreted the telling of a narrative conveyed through speech accompanied by complementary, nonredundant gestures. Based on participants' recollections, it was found that information from gesture is incorporated into listeners' overall understanding of a story. In addition, when participants were asked to retell the story, information they received through gesture was oftentimes realized verbally. These results further indicate that meaning expressed in the two modalities is processed together as a single, unified whole.

Research in neuroscience corroborates these findings. Wu & Coulson (2007) measured ERPs while participants observed spontaneously produced speech-and-gesture utterances from a conversation. It was found that semantic activations from viewing gesture are used in parallel with those from speech, and the authors claim the two are processed together to jointly build conceptual representations during discourse comprehension. Others have found similar results indicating comprehension involves the simultaneous integration of the modalities (Cornejo, Simonetti, Aldunate, Ceric, Lopez, & Nunez, 2009; Ozyurek et al., 2007). In a more fine-grained analysis, Kelly, Kravitz & Hopkins (2004) further demonstrated the connections between gestural and speech comprehension by showing that gestures are processed like words at both early and late stages of semantic processing, indicating the coordination of speech and gesture occurs at a deep semantic level.

In addition to this evidence that gesture and vocal language are processed together, other studies have found that gesture actually improves memory for the accompanied speech. Feyerseisen (2006) tested participants' recall of sentences presented as speech-only or as speech with corresponding gesture. Sentences with gestures were correctly recalled significantly more often than their speech-only counterparts. Kelly et al. (1999) also found that recall of words from speech is significantly better when participants are presented with speech-plus-gesture utterances than speech alone. Beattie & Shovelton (1999) found overall accuracy for recalling stories is greater when

narration is presented through speech and gesture as opposed to speech alone. In a somewhat different context, Kelly, McDevitt & Esch (2009) tested learning of new words in a foreign language. They found the highest rates of word learning occurred when participants were trained on words presented with representational gestures, leading the authors to suggest that distributing semantic content between the two modalities enhances depth and strength of memory. It is important to note that the training regime consisted of presenting new foreign words together with *both* a gesture and an English translation. That is, the learning task involved tapping into participants pre-established symbolic representations.

The reviewed research indicates that information from the gestural and vocal modalities is instantly and effortlessly combined and interpreted as a single unit. In fact, in some cases the presence of gesture improves comprehension and memory for meaning expressed in speech. The reason for this integration may simply be the deep anatomical and psychological connections between the two modalities (discussed in Chapter 3 section 3.2), which supports the proposal that vocalizations perceived as arbitrary would be attended to and potentially reinterpreted as symbols. However, another possibility that cannot be excluded based on this evidence is that these findings result from pre-established connections between vocal words and gestures; each carries meaning related to the other, and additional sources of evidence indicate that this relationship is reflected in the two sharing a common representational.

Research on the brain and language has demonstrated the direct involvement of sensorimotor representations in language processing. Masson et al. (2008) observed brain activity while participants read sentences with words referring to manipulable objects and found that these words evoke the same neural activations for specific, appropriate hand movements for the corresponding objects. Boulenger et al. (2006) tested how language influenced reaching movements and showed that presentation of action verbs facilitated reaching. Moreover, the effects were body part specific – in other words, facilitation was greater with words for actions involving the body part employed in the reaching task. Rueschemeyer et al. (2010) demonstrated that this

effect is bidirectional, finding that hand movements facilitated the processing of semantically related words for manipulable objects. Glenberg & Kaschak (2002) also found a language and sensorimotor influence, but in this case an interference was observed when the direction in which participants were required to move conflicted with the direction conveyed in a sentence just read. A number of other neuroscientific studies have demonstrated that words and sentences referring to body parts or actions activate areas of the motor cortex specific to those parts and associated actions (Aziz-Zadeh et al., 2006; Hauk et al., 2004; Pulvermuller et al., 2005; Tettamanti et al., 2005; Willems et al., 2010). Finally, Parrill, Bullen & Hoburg (2010) found no difference in the gestures produced in spontaneous narratives by participants who read a story in text versus those who watched a video, leading them to conclude that language comprehension involves simulation of sensorimotor representations to construct imagined perceptual experiences. While these examples for the most part concern concrete, action-related words, there is also some evidence that the motor system is in part responsible for representing and processing abstract language. In Masson et al.'s (2008) experiment, object-specific manual actions were evoked for sentences that contained no action words - for example, 'John thought about the calculator' (pg. 878). In addition, Glenberg et al. (2008) found that processing concrete and abstract sentences modulated the motor system in similar ways. Though findings are not yet definitive, the overall picture suggests the sensorimotor system is at the very least involved in language processing generally and not solely for action-related words.

Given that gestures are rooted in sensorimotor representations, we would expect the connection between vocal language and the sensorimotor system to extend to gestures, and some studies indicate this is the case. Xu, Gannon, Emmorey, Smith & Braun (2009) found that comprehension of spoken language and gesture depends on a shared network of overlapping brain regions. Hubbard, Wilson, Callan & Dapretto (2009) also showed that speech and gesture processing share a neural substrate. Finally, and regarding abstract language, Cornejo et al. (2009) showed that gestural information is used in the comprehension of metaphorical expressions in speech. Thus, words and gestures with related meanings appear to derive from overlapping representations, and

it could be that this common representational origin is responsible for the integrated and improved processing observed in the gesture studies described above.

In order to determine whether vocalizations in the pre-symbolic context described in the previous chapter would be processed like words and language today, and therefore potentially reinterpreted as symbols, we would need to demonstrate similar results without the explicit involvement of participants' established linguistic system. One way to do this would be to measure if and to what degree new vocalizations are learned, which serves as an indicator of whether participants are attending to and potentially adopting vocalizations as meaningful symbols. In addition, the task should not be an explicit learning task, like that in Kelly et al. (2009), as we must assume no expectation to learn a symbolic system on the part of communicators. As such, our experiment can be a task in which participants attempt to guess the meaning of unfamiliar vocalizations. After guessing, participants will be informed of the correct meaning, providing the opportunity for reinterpretation and learning. In addition, to simulate the proposed misinterpretation-to-reinterpretation process outlined in the previous chapter, we would want participants to expect vocalizations to be nonarbitrary and base their guesses first on a search for meaning in vocal forms. By using vocalizations designed not to depict nonarbitrary features of their referents, we can expect participants will not successfully guess meanings and thereby interpret vocalizations as arbitrarily related to those meanings. Finally, to demonstrate that acquiring meaning through a coordinated nonarbitrary gesture does not overshadow and remove vocalizations from interpretive processes, we can test whether presentation via multimodal utterances also results in vocalizations being learned and whether learning differs between the two. The following section describes an experiment designed to address these questions.

3. Experiment I

This experiment will be a two-part task in which participants first watch videos of a person saying unfamiliar 'words' and attempt to guess their meaning. Following that,

participants will be presented with the words again and asked to identify their correct meanings. The experiment will have to conditions: vocal-only (VO) and vocal-plus-gesture (VG). In the VO condition, words will be spoken with no accompanying gesture, and after guessing participants will be shown the correct meaning. In the VG condition, words will be accompanied by a related nonarbitrary gesture, and afterward participants will be still the shown the correct meaning. In the VG condition, we expect participants to interpret the meaning of the word based on the gesture and therefore guess correctly. In all cases, participants have access to meanings and therefore the potential to learn the words. In the VO condition, we expect participants to guess incorrectly and interpret the meaning of the word based on the feedback they receive after guessing.

We expect that if we test participants' memory for VO words, they will demonstrate learning. In the VG condition, participants have access to meaning through gestures and do not need to attend to or utilize features of the vocalization in order to guess correctly. In addition, it may be that the movements alone could be distracting. Therefore, one possibility is that participants in this condition will ignore or lose focus on words. If this were the case, participants may fail to learn words at all or to a lesser extent than the VO condition. Alternatively, if the words and semantically related gestures are perceived and processed as an integrated unit, learning patterns may be similar to the VO condition.

The claims made in this thesis regarding the interpretation of novel arbitrary vocalizations predict certain results. First, learning will occur in both the VO and VG conditions. An additional expectation that follows from integrated processing is that the extent of learning – or the number of words remembered – will not be negatively impacted by gesture. Thus, I predict that performance in the VO condition will not be higher than the VG condition. The specific hypotheses we are testing are as follows:

i) Words will be learned in the VO condition
ii) Words will be learned in the VG condition
iii) The number of words learned in the VO condition will not be significantly higher than the VG condition

If i) is correct, it suggests that novel arbitrary vocalizations would be interpreted as meaningful and potentially adopted as symbols. If ii) and iii) are correct, it indicates integrated processing of meaningful representations in the two modalities occurs even in the absence of a pre-established representational connection. By extension, that would suggest that communication through nonarbitrary gestures would not impede the interpretation of vocalizations as symbolic and therefore could also have functioned as an important grounding source for the establishment and conventionalization of novel vocal symbols.

Finally, it is necessary to clarify how learning will be measured in order to test these hypotheses. The type of learning involved in this task will be implicit, as participants are merely guessing word meanings and have no expectation to acquire or use the words. Choosing a specific number or proportion of words remembered as a threshold for learning is somewhat arbitrary and not a very useful measure. Instead, learning will be measured by deviance from chance. To make this a more robust indicator, our criterion for learning will be a highly significant deviation from chance – a p-value below .001. If participants' performance in correctly identifying meanings reaches this level, it indicates they have attended to and learned the novel arbitrary vocalizations.[12]

3.1 Methods

3.1.1 Design

The experiment was conducted using a one-way, within-subjects design comprising two separate tasks: a Guessing portion followed by a Testing portion.

3.1.2 Materials

The stimuli consisted of 40 words and 160 images of objects with familiar associated actions. Images were used as meanings instead of translated English text to avoid explicitly activating linguistic, symbolic representations. Objects were chosen from an inventory of 480 normative stimuli developed by Brodeur, Dionne-Dostie, Montreuil, Lepage & Op de Beeck (2010). All objects had normativity ratings at or

[12] Whether these are in fact arbitrary is also important to establish. How this was checked is discussed in section 3.2.

above 3.5 (out of a maximum of 5) for 'manipulability', defined as how easily one could produce a gesture to represent the object. 40 of these objects were randomly assigned as the 'correct' meaning for each of the 40 words. This was done i) so that nonarbitrary gestures could be created and correctly perceived by participants, and ii) to ensure the objects chosen to be gestured were not necessarily the ones with highest manipulability, as this in itself might affect the results. The remaining objects were randomly assigned to the 40 words to serve as three additional potential meaning choices.[13] Words were of Japanese origin and paired with objects randomly in order to reduce the possibility that their forms reflected a nonarbitrary relationship to their intended meanings. Japanese was chosen because its phonological inventory and phonotactics involve distinctions found in many languages and could therefore be perceived by speakers of a variety of languages.

A volunteer speaker was recorded saying the 40 words two times for each condition using a Kodak Zi8 digital video camera and AKG CK98 hypercardioid microphone with an Alice MIC AMP PAK II preamp. For VO, the speaker spoke standing with his arms at his sides, facing the camera. For VG, the speaker said the word while also producing an iconic or icon-index hybrid gesture depicting an action performed on or with the object. These were edited into 80 4.5 second video clips and used for the guessing portion. Figure 4.1 shows stills from videos for the two conditions. With the same microphone, the speaker was audio recorded saying the 40 words in a sitting position, which were edited into 40 2.5 second clips and used for the testing portion. This was done to ensure the audio used in testing did not resemble the words spoken from any one condition more than the other.

3.1.3 Procedure

Participants were told that the experiment was designed to test how well people can guess the meaning of words in a foreign language (only participants with no experience with Japanese were included). Specifically, they were told it was based on the concept of onomatopoeia and that words somehow 'sound like' their meanings. This was done to create the expectation for nonarbitrariness and encourage

[13] Random assignments were checked for the possibility that the gesture for a correct object could be confused with one of the 'incorrect' objects, and in a small number of cases objects were reassigned.

participants to utilize this strategy. Participants were told the experiment included a second round that would be similar to the first, but they had no knowledge that their memory for the words would be tested. After completing the guessing portion, participants were informed that they would now be tested to see how well they remembered the meaning of the words they had just guessed. The experiment lasted approximately twenty minutes and each participant received three pounds sterling.

a.

b.

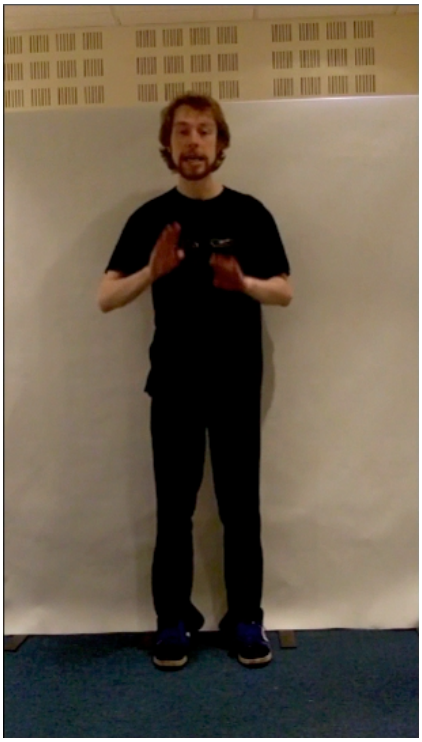


Figure 4.1. Stills from videos for the word meaning 'drum' in the VO condition (**a**) and VG condition (**b**)

The experiment was presented through an E-Prime program on PCs running Windows 7. It consisted of a series of video clips presented with four unique images of objects – that is, objects that appeared with one word did not appear with any other words. Figure 4.2 depicts the layout of a single guessing screen. After a video played, participants were instructed to click on the object they thought the word meant. Once a guess was made, all objects but the correct one disappeared from the screen, and text appeared informing the participant that the remaining object was the correct meaning. Participants could then press the space bar to move on to the next video

clip. When the guessing portion was completed, the experimenter was called in and explained the testing section. Participants were also presented with written instructions and training for this section. Testing consisted of a series of audio clips presented with four images of objects. The set of objects for each word included 1) the correct object, 2) an object that appeared with that word in the guessing task but was *not* the correct meaning, 3) an object that did not appear with the word in the guessing task but *was* the correct meaning for a *different* word and 4) an object that appeared with a *different* word (not the same word as 3) but was *not* the correct meaning. Participants did not receive feedback in testing in order to preclude using a process of elimination strategy over the course of the block. After completing the testing portion, participants were asked questions about the experiment regarding their confidence in remembering words, whether they noticed the speaker gesturing and how it affected their guessing, and whether they thought the words 'sounded like' their meanings.

Two versions of the program were created corresponding to the two experimental conditions. Each version included twenty words each of VO and VG video clips, such that each participant saw all 40 words, half for each condition, and all words for both conditions were included overall across participants. The order of words in the guessing portion was randomized one time, and all participants experienced the task in this same order. Each version had two sub-versions in which the order that words were presented in testing was different.

Click on the object that you think the word means from the set below.

Figure 4.2. Layout of screen after video has played and participants guess which object the word they just heard means.

3.2 Results

A total of 41 participants were recruited, 33 female, eight male.

As expected, words in the VG condition were generally guessed correctly, with performance highly significantly greater than chance (one sample t-test, $M = 18.83$, $SD = 1.38$), $t(40) = 64.32$, $p < .001$, $d = 13.83$). Also as expected, guessing performance on average in the VO condition did not rise above chance, and in fact

was significantly below chance (one sample t-test, $M = 4.26$, $SD = 1.5$, $t(40) = -3.12$, $p = .003$, $d = -0.73$).

As predicted, the number of words remembered in the testing phase deviated from chance highly significantly for both the VO (one sample t-test, $M = 13.31$, $SD = 3.01$, $t(40) = 17.68$, $p < .001$, $d = 8.32$) and VG ($M = 13.80$, $SD = 2.27$, $t(40) = 24.82$, $p < .001$, $d = 8.80$) conditions. These chance figures are based on the reasoning that there were 4 options to choose from in any given case for all 13 words in each condition. As described above, however, the testing phase options included only two objects that were 'correct' answers previously, a fact that might be recognized by participants. It is therefore arguable that a more accurate standard for chance in this case may be based on there being two options instead of four. By this measure, VO (one sample t-test, $t(40) = 7.05$, $p < .001$, $d = 3.32$) and VG ($t(40) = 10.72$, $p < .001$, $d = 3.80$) conditions remain highly significantly above chance.

In comparing the two conditions, on average, more VG words were remembered than VO words, though this difference was nonsignificant (two-tailed paired sample t-test, $t(40) = -.915$, $p = .366$, $d = -.975$). Correlations were run to check whether words that were guessed correctly were more likely to be remembered in testing, which is one possible explanation for why more VG words were remembered than VO words. However, no significant correlation was found overall (Spearman's rho = .145, $p = .200$), for VG words alone (Spearman's rho = .227, $p = .160$) or for VO words alone (Spearman's rho = .141, $p = .386$).

Exit questions revealed that participants generally rated their memory for words fairly low (average 4.2 on a scale form 1-10). All reported noticing the speaker gesturing and that they were usually able to guess based on the gesture. Regarding the vocalizations themselves, the overwhelming majority of participants thought that the words did not sound like their meanings at all. Some reported that certain individual words showed resemblance to their meaning, and these included the words *chigau* ('lipstick'), *hidari* ('toothbrush') and *soko* ('bracelet'). Notably, none of these was cited by more than one person, indicating that, overall, vocalizations were generally perceived as arbitrary. In addition, with the exception of *chigau* in the VO condition, these words were not remembered on average above the mean for all words,

indicating words that may have been more likely to be perceived as nonarbitrary were not more likely to be learned than others.

3.3 Discussion

These results support hypotheses i) – iii) laid out above. By strict measures against chance, participants learned words in both conditions. In addition, VG words were not remembered to a lesser extent than VO words, demonstrating that the gestures did not disrupt or obviate the interpretation and learning of the words.

Importantly, answers given to exit questions indicate that, as intended, the words were perceived as arbitrary. Given that participants initially attempted to understand the vocalizations as nonarbitrary representations, we can reason that their learning of words involved a process of misinterpretation-to-reinterpretation.

The finding that VO and VG were learned in these experimental circumstances indicates that, even when there is not expectation to learn a symbolic vocabulary and a listener is attempting to understand a message through recognition of nonarbitrary features, novel arbitrary vocalizations are attended to and remembered. This suggests that arbitrary vocalizations would have been perceived as symbols and potentially learned and used later. In addition, the memory for VO words was not greater than VG words, with no significant difference observed between the two conditions. Thus, nonarbitrary gestures that carry the full communicative load do not detract from accompanying vocalizations and impede the processing and learning of those vocalizations, indicating that signs in the two modalities are integrated even when those do not share an established association via a common representational source.

These results show the positive result that words in both conditions are learned. In a comparison between conditions, however, the results only show that there is *not* a decrease in learning between VO and VG words. A weakness in this experiment, then, is that it demonstrates a negative result. Stronger evidence for the relationship between signs in the two modalities may be obtainable by comparing the effects of other gestures on learning in the same task. Speech is often accompanied by movements that are not communicative signs, and testing to see if these affect the

processing and interpretation of vocal words in a different way may highlight the unique relationship between gestural and vocal signs. Thus, we can investigate the effects of multimodality specifically as we are interested in it here by testing the effects of non-communicative gestures frequently produced during communication. So-called 'self adaptors' are movements that accompany speech that are not intended to represent or explicitly communicate information related to the content of speech. These include actions like tucking hair behind the ear, adjusting clothes or glasses, rubbing the nose and scratching. In the same way that meaningful gestures could potentially distract attention from vocalizations, movements of this sort could do the same but not provide a communicative representation related to a speaker's message. Testing the effect of these gestures will allow us to better understand how gestural *signs* specifically affect interpretive processes. The next section will describe a second experiment designed to address these questions.

4. Experiment II

This experiment will be similar in design to the preceding one and compare learning under two conditions. In this case, the conditions will contrast vocal-only (VO) words with vocal-plus-adaptor (VA) words. VA words will be accompanied by self-adaptor gestures and followed by feedback. In the VO condition, we again expect participants to guess incorrectly and interpret the meaning of the word based on the feedback they receive after guessing. As with the preceding experiment, we expect that if we test participants' memory for these words, they will demonstrate learning. In the VA condition, we again expect participants to guess incorrectly and interpret the words based on feedback. We would still expect VA words to be learned; however, it may be that the accompanying movements are distracting and result in diminished learning. If self-adaptor gestures have such an effect, in contrast to nonarbitrary gestures, we would expect learning for VA words to be significantly lower than VO words. Thus, the specific hypotheses we are testing are:

i) Words will be learned in the VO condition

ii) Words will be learned in the VA condition

iii) The number of words learned in the VA condition will be significantly lower than the VO condition

If i) and ii) are correct, it further supports findings from the previous experiment and the arguments presented in this thesis regarding the interpretation of novel vocal symbols. If iii) is correct, it demonstrates that gestures can interfere with interpretive processes. It would further indicate that the results above (section 3.2) are specifically attributable to the particular properties of nonarbitrary gestural signs and their unique relationship to the vocal signs they accompany.

4.1 Methods

4.1.1 Design

The experiment was conducted using the same one-way, within-subjects design comprising two separate Guessing and Testing tasks.

4.1.2 Materials

The stimuli consisted of the same 40 words and 160 images of objects. The same word-meaning pairings were also used.

New videos for each condition were created, with the volunteer speaker saying the 40 words two times for each condition. For VA, the speaker said the word and also produced a self-adaptor gesture. These were edited into 80 4.5-second video clips and used for the guessing portion. Figure 4.3 shows stills from videos for these two conditions. The same audio clips from the previous experiment were used for the testing portion.

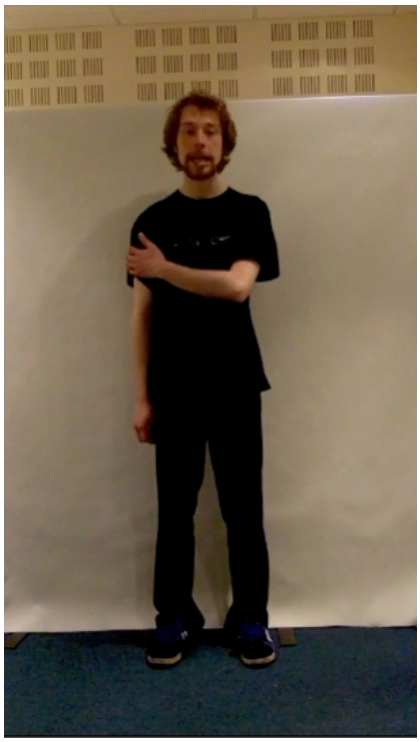
a.  b. 

Figure 4.3. Stills from videos for the word meaning 'drum' in the VO condition (**a**) and VA condition (**b**)

4.1.3 Procedure

Participants received the same instructions regarding the task, and the same E-prime program was used. The structure of the experiment remained like the previous, with ordering of stimuli, program versions and other details unaltered (see section 3.1.3 for a full description).

4.2 Results

A total of 25 participants were recruited, 13 female, 12 male. One male was excluded due to participant error[14], resulting in 24 datasets included in analysis.

As expected, guessing performance in both VO (one sample t-test, M = 4.96, SD = 1.71, $t(23)$ = -0.12, p = .906, d = -0.04) and VA (M = 4.75, SD = 1.94, $t(23)$ = -0.63, p = .534, d = -0.25) conditions did not deviate from chance significantly. On

[14] During exit questioning it became clear he knew Japanese, which was cause for exclusion.

average, fewer words were guessed correctly in the VA condition than the VO condition, but this difference was not significant (two-tailed paired t-test, $t(23) = 0.359$, $p = .723$, $d = 0.417$).

As predicted, the number of words remembered in the testing phase for both VO (25%: $M = 13.5$, $SD = 2.92$, $t(23) = 14.26$, $p < .001$, $d = 8.5$; 50%: $t(23) = 5.87$, $p < .001$, $d = 3.5$) and VA (25%: $M = 12.46$, $SD = 2.40$, $t(23) = 15.26$, $p < .001$, $d = 7.46$; 50%: $t(23) = 5.03$, $p < .001$, $d = 2.46$) conditions was highly significantly above chance, for both a 25% and 50% baseline.

In comparing the two conditions, participants, on average, remembered more VO words than VA words. A t-test revealed this to be a significant difference (one-tailed paired t-test, $t(23) = 1.78$, $p = .044$, $d = 2.08$) (a one-tailed test was used as hypothesis iii predicts a particular outcome – that more VO words will be learned than VA words).

Answers to exit questions were similar to those given in the previous experiment. The mean confidence rating for remembering words correctly was 4.8. Words were again generally judged as not resembling their meanings. The few words reported to sound like their meanings by participants in Experiment I were not cited in this experiment, though *hoshi* (match) was by two participants and *kawaii* (drum) by one. Participants frequently described the speaker's gestures as distracting or off-putting, with some even speculating that the speaker was intentionally trying to mislead their guesses. Some participants claimed to have simply ignored the gestures altogether, though this did not appear to translate into greater than average performance (of the two participants who made such claims, one remembered eleven VO words and the other fourteen, compared to the mean 12.45).

4.3 Discussion

These results support the hypotheses i) – iii) above. As in the previous experiment, words were learned in both conditions. However, unlike the comparison between vocalizations alone and vocalizations with related nonarbitrary gestures, vocalizations accompanied by self-adaptor gestures were remembered significantly less than

vocalizations only. Answers to exit questions again suggest that the words were perceived as arbitrary as intended, and a similar process of misinterpretation-to-reinterpretation occurred.

The finding that words in both conditions were learned corroborates the previous experiment's results and further supports claims made in this thesis regarding novel vocal symbols. In addition, the significant reduction in learning observed for VA words in comparison to VO words indicates that movements occurring with vocalizations can in fact disrupt memory and interpretation processes. The effect that self-adaptors have reveals the importance of the lack of such an effect for nonarbitrary gestures. Interestingly, adaptor gestures were generally shorter and smaller than nonarbitrary gestures used in the previous experiment, which on the surface suggests adaptors would be less distracting. Nevertheless, they impeded word learning compared to no gesture while nonarbitrary gestures did not. The divergent effects of non-communicative gestures versus gestural *signs* highlight the psychological unification of intentional multimodal communication and that interpretive processes naturally integrate semantically related representations in the two modalities. However, a problem with this reasoning that we cannot directly compare the present results to those from the preceding experiment. To fully substantiate my claims based on these results requires testing all three conditions at once. Doing so will allow us to test the effects of gestures in comparison to no gestures, as well as to compare the two types of gesture to one another.

Another potential issue with this experiment is the strong reactions participants had to self-adaptors. Given that gestures of this sort frequently occur during communication, these reactions are at first surprising. However, observations from research on language and gesture point to a possible reason. A hallmark of nonarbitrary co-speech gestures is precise temporal and semantic alignment with concurrent words; that is, the movements constituting representations in gesture are tightly linked to vocal sounds and energy patterns (Chapter 3, section 3.2). Self-adaptor gestures are not related to the content of speech and typically do not show such precise alignment. A close inspection of the speaker's actions in the video clips used here found that, unlike in normal communication, adaptors were in tight temporal alignment with

spoken words.[15] This atypical alignment could in fact be an added distraction, and the observed effect of VA words may disappear if the speaker's behavior were adjusted to resemble normal behavior. To ensure the findings here are not a consequence of this possible confound, another experiment must be conducted using videos in which the speaker's gestures more closely match natural self-adaptors. The following section describes an experiment that combines VO, VG and VA conditions with corrected stimuli.

5. Experiment III

This experiment will be similar in design to the preceding one, except learning will be compared between three conditions: vocal-only (VO), vocal-plus-nonarbitrary gesture (VG) and vocal-plus-adaptor (VA). As with prior versions, we expect words to be learned in all three conditions. Based on the results of Experiment 1, we expect to find no difference between learning in VO and VG conditions. In this case, we want to go further distinguish between the two gesture conditions. Based on results from both experiments, we expect performance in the VA condition to be significantly lower than the VG condition. Thus, the hypotheses are:

> i) Words will be learned in all conditions
> ii) The number of words learned in the VG condition will not be significantly lower than the VO condition
> iii) The number of words learned in the VA condition will be significantly lower than the VG condition

If i) is correct, it supports the findings from Experiments I and II. If ii) is correct, it supports results from Experiment I and arguments based on them (section 3). If iii) is correct, it will demonstrate the distinctive properties of nonarbitrary communicative gestures and their unique relationship to the vocal modality.

[15] The reason for this probably lies in the fact that the speaker was consciously aware of his actions, and self-adaptor gestures are generally not produced deliberately with overt attention.

5.1 Methods

5.1.1 Design

The design was the same as that Experiments I and II, this time with three conditions.

5.1.2 Materials

The stimuli consisted of 39 words of the 40 used in previous versions (for three equal proportions of words for each condition) and 156 of the 160 images of objects. The one word removed was selected randomly, and all word-object pairings were the same as Experiments I and II.

A speaker was video recorded saying the 39 words three times each for each condition. For VO, the speaker spoke standing with his arms at his sides, facing the camera. For VA, the speaker said the word and also produced a self-adaptor gesture – this time carefully timed not to coincide perfectly with energy peaks in the vocal modality.[16] These were edited into 117 4.5-second video clips and used for the guessing portion. Figure 4.4 shows stills from videos for all three conditions. The same audio clips from previous experiments were used for the testing portion.

[16] Movements were still concurrent and overlapping with vocal sounds.

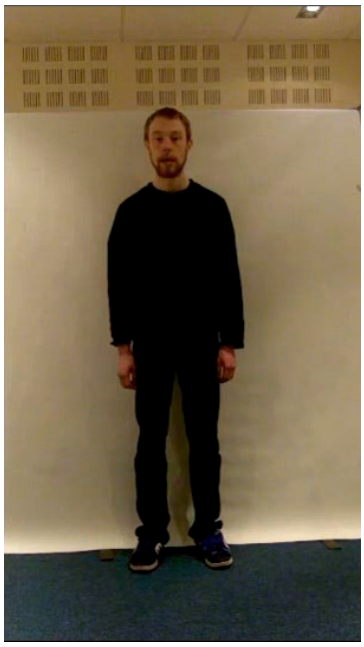
a.  b.  c. 

Figure 4.4. Stills from videos for the word meaning 'drum' in the VO condition (**a**), VG condition (**b**) and VA condition (**c**).

5.1.3 Procedure

Participants received the same instructions regarding the task, and the same E-prime program was used. Two versions of the program were created corresponding to the two experimental conditions. Each version included 13 words each of VO, VG and VA video clips, such that each participant saw all 39 words, one-third for each condition, and all words for all three conditions were included overall. The structure of the experiment otherwise remained like the previous, with ordering, sub-versions and other details unaltered (see section 3.1)

5.2 Results

A total of 30 participants were recruited, 17 female, 13 male. One male was excluded due to participant error, resulting in 29 datasets included in analysis.

As expected, words in the VG condition were generally guessed correctly, with performance highly significantly greater than chance (one sample t-test, M = 12.55, SD = 0.63, $t(28)$ = 79.30, p = < .001, d = 9.30). Also as expected, guessing

performance in both VO (one sample t-test, M = 3.21, SD = 1.26, $t(28)$ = -0.18, p = .855, d = -0.04) and VA (M = 3.69, SD = 1.61, $t(28)$ = 1.47, p = .152, d = 0.44) conditions did not deviate from chance significantly.

For a 25% baseline, the number of words remembered in the testing phase deviated from chance highly significantly for all three VO (one sample t-test, M = 8.31, SD = 2.27, $t(28)$ = 12.01, $p < .001$, d = 5.06), VG (M = 8.97, SD = 1.97, $t(28)$ = 15.60, $p < .001$, d = 5.72) and VA (M = 7.45, SD = 1.88, $t(28)$ = 12.01, $p < .001$, d = 4.20) conditions. With a 50% baseline, VO (one sample t-test, $t(28)$ = 4.30, $p < .001$, d = 1.81) and VG ($t(28)$ = 6.73, $p < .001$, d = 2.47) conditions remained at the same level of deviation. VA words were at a lower level of significance, though still well above chance ($t(28)$ = 2.71, p = .011, d = 0.95).

On average, participants remembered more VG words than VO words and more VO words than VA words. A one-way repeated measures ANOVA was run and returned a significant effect of condition ($F(2, 56)$ = 6.51, p = .003). Planned contrasts showed the number of VO words remembered approached but did not reach significance in comparison to VA words ($F(1, 28)$ = 3.74, p = .063). Post-hoc analyses using a Bonferroni adjustment for multiple comparisons revealed no significant difference between VO and VG conditions (p = .492), but significantly more VG words were remembered than VA words (p = .001).

Correlations were run again to check whether words that were guessed correctly were more likely to be remembered in testing, which is one possible explanation for why more VG words were remembered than other groups. The mean number of times a word was remembered was 6.18 for the VO condition, 6.67 for VG and 5.54 for VA. The mean number of correct guesses for words was 3.21 for the VO condition, 12.55 for VG and 3.69 for VA. However, no significant correlation was found (Spearman's rho = .116, p = .215).

Answers to exit questions were similar to Experiments I and II. The mean confidence rating for remembering words was 3.3, and participants in large part reported that words did not sound like their meanings. Two exceptions were champagne (*yappari*)

and hairclip (*simasu*), neither of which was cited by more than one participant in this experiment or by participants in previous experiments.

5.3 Discussion

These results corroborate findings from Experiments I and II and support the specific hypotheses laid out above. Participants learned words in all three conditions, though based on our stricter criterion VA words were learned to a slightly lesser degree than in Experiment II. In addition, VO and VG conditions showed similar learning levels, while the VA condition had distinct effects from the VG and verged on doing so for the VA condition. That the VA condition still significantly reduced learning after correcting for the possible confound in Experiment II demonstrates that effects are due to self-adaptors themselves and not atypical gestural timing. Finally, exit questions again revealed that, as intended, the words were perceived as arbitrary, indicating the desired misinterpretation-to-reinterpretation process took place for some words.

The finding that VO and VG were learned in these experimental circumstances is further indication that novel arbitrary vocalizations are processed similarly to novel symbolic words. In addition, no difference was found between these two conditions, mirroring results from Experiment I and again indicating that signs in the two modalities are integrated even when those do not share an established association via a common representational source.

The observation that learning levels are similar for VO and VG conditions becomes striking when we consider the effect of the VA condition. Noncommunicative gestures impeded word learning compared to no gestures while nonarbitrary gestures did not, despite nonarbitrary gestures being larger and longer, and therefore potentially more distracting. Instead, VO and VG conditions patterned together, demonstrating that nonarbitrary gestures have a unique relationship to the vocal signs they accompany, and confirming conclusions from Experiments I and II that vocalizations are interpreted and processed in similar ways for these conditions.

VA words were also learned in this experiment, though they showed less robust learning under the more stringent measurement of learning. This is in contrast with the results of Experiment II, which demonstrated learning for VA words by both standards. These different learning levels across experiments may have to do with the presence of multiple gesture types. In this case, gestures could be either informative or distracting, but gesture type would be inconsistent throughout the task. The fact that participants could not wholly rely upon or ignore gestures may have negatively impacted learning overall. This seems to be the case, as the proportion of words learned dropped to 62% in Experiment III, compared to 68% in Experiment I and 65% in Experiment II. Interestingly, however, the decrease only occurs for VA words (62% in II to 57% in III) and VO words (68% and 67% in I and II to 64% in III), with the proportion of VG words learned holding constant at 69% from Experiments I to III. This suggests that inconsistency in expression may not negatively impact learning when nonarbitrary gestures *are* produced, but learning words under other conditions will be.

6. General discussion

Findings from Experiments I-III suggest that, in the circumstances of presymbolic communication as described in Chapter 3, novel arbitrary vocalizations would have been perceived as symbols, learned and potentially used later. Even in the absence of an expectation to acquire a symbolic system and with communication dominated by nonarbitrary gesture, these results suggest that vocalizations would be perceived as intentional, meaningful signs on par with gesture and, when perceived as arbitrary, interpreted as symbolic signs and learned similarly to how symbolic words are today. In this way, vocalizations were interpreted as arbitrary and learned without any of the communicative pressures known to favor symbolic signs (Chapter 1, section 3.3). Thus, overall these findings support the proposal in this thesis that vocal symbols could naturally arise from interpretive processes in operation during (intended) homeosemiotic multimodal communication.

It is not possible to directly compare the extent of learning in these experiments to that observed in Kelly et al.'s (2009) study on foreign language learning reviewed in

section 2 above. Their study included a total inventory of only twelve words and participants were exposed to these a total of three times, as opposed to the single time in the present experiments. Nevertheless, it is interesting to note that in a test of recognition wherein participants had to choose the correct meaning of a word from a set of four choices, on average, words were remembered 78% of the time. Averaged across Experiments I-III here, that percentage is 65%, which could arguably be considered on par to their findings, adjusting for the more than tripled inventory size, more limited exposure and no intention to acquire new words. This similar pattern of learning to a task in which participants are explicitly learning new symbols further suggests interpretations of vocalizations as symbols could arise without additional pressures from the conditions of communication described in Chapter 3 section 6.2.

Present results also demonstrate the integrated nature of multimodal communication as observed in the studies on language and gesture reviewed in section 2. An additional but less robust finding from both Experiments I and III is that words accompanied by nonarbitrary gestures were learned to a greater degree than words alone. This pattern reflects evidence from other studies that demonstrate memory for speech is improved when accompanied by a semantically related gesture (Beattie & Shovelton 1999; Feyerseisen, 2006; Kelly et al. 1999; Kelly et al. 2009). The increase in memory was not significant in the present experiment, but these results remain inconclusive. One possibility is that enhanced memory effects result from existing associations between words and gestures by way of shared representational sources. That is, when two, overlapping representations are activated for particular semantic content – in this case a spoken symbol and a related nonarbitrary gesture – these could reinforce each other and thereby increase memory for one or both. In the present experiment, such reinforcement would not occur, as the words participants experience do not yet carry any meaning for them.[17] Thus, it is possible reinforcement via dual representations is responsible for increased memory. Alternatively, it may be that the lack of significance could be an artifact of the experimental design, and gesture could have a significant effect in slightly altered circumstances. For instance, in the

[17] It is of course possible that participants think of a word in their language when guessing the novel words' meaning and receiving feedback, but a specific word is not explicitly brought to attention as in other experiments. It is the specific, overt involvement of established linguistic symbols that might be influencing results.

experiment most similar to the present one, Kelly et al.'s (2009) foreign language learning paradigm, the design was between subjects, which means an individual participant only experienced words in a single condition. In this case, the presence of gesture would be consistent and expected, and participants could learn to attend to and utilize it as a reliable source of information. In contrast, participants in this experiment only experienced nonarbitrary gestures for one third of words. Sometimes gestures were not present at all, sometimes gestures were present but gave no clue to the intended meaning. The inconsistency inherent to a within-subjects design might diminish the memory boost provided by gesture and could explain the difference in results between this and Kelly et al.'s (2009) study. If so, gesture may have the added effect of increasing learning of new vocal symbols in the context of emergence. Improved learning of this sort could itself have further effects on conventionalization processes, particularly when we consider the additional finding that inconsistency in expression reduced learning rates for words alone and with noncommunicative gestures (section 4). By aiding memory for novel symbols, gesture could have provided additional support and momentum for their spread and adoption by others. In addition, it could also affect the extent that arbitrary signs are conventionalized across a population and may be instrumental in explaining the observed differences in conventionalization in speech compared to signed languages. These are questions that can be explored in future research.

7. Conclusion

Previous research demonstrates that signs in speech and gesture are interpreted and processed together as a unified whole. The experiments here show that this integration holds even when the communicative roles of the modalities are asymmetrical and no prior connection exists between vocal and gestural content. These results support the claims made in the previous chapter and inform our understanding of the interaction of the modalities in sign interpretation processes. The results also raise questions and point to future research avenues regarding the possibility that gesture in fact increases learning, which, if true, opens the possibility that gesture could have further effects on the process of conventionalization. The next chapter will address the issue of conventionalization and the scenario proposed in Chapter 3 section 6.2.

CHAPTER 5
Conventionalization: Computational simulations

1. Introduction

A theme of this thesis is that a consideration of the fundamentally multimodal nature of embodied human communication can inform and aid an explanation of symbol origins. The preceding chapter addressed questions concerning interpretation, and this chapter will address conventionalization – how, once arbitrary interpretations of signs arise, populations come to share the same symbols.

In Chapter 3, section 6.2.2 I argued specifically that nonarbitrary gestures would have provided sufficient grounding for arbitrary vocalizations, which in turn would enable conventionalization. In terms of mechanisms underlying grounding, this implies that information provided by gesture could support joint attention and reduce the context of communication sufficiently for learning and cultural processes to lead to convergence on a shared symbol system. This chapter will approach this issue as it has been traditionally, through computational simulations. Previous research on the topic has been reviewed and discussed in detail in Chapter 2, section 4. This chapter will therefore first briefly revisit how existing models have been structured and how this framework can be modified to incorporate multimodality and different sign types. Three models will then be presented that are designed to test the claims made in this thesis regarding nonarbitrary gesture and conventionalization.

2. Background

Computational models of symbol emergence begin with the assumption that signs are arbitrary. Broadly, models simulate a population of agents that send and receive messages containing arbitrary signals, and through repeated exchanges agents learn signal-mappings. Conventionalization occurs when agents in a population come to share – or converge on – the same signal-meaning mappings. Successful convergence relies on a mechanism to simulate joint attention, which serves to limit the context of communication and allow receivers to identify a senders' intended meaning. Some

models implement perfect context reduction, wherein agents receive only a single meaning with signals and are therefore able to learn the 'correct' mapping (Steels & Kaplan, 1998; Steels et al., 2001; K. Smith, 2002). Steels & Kaplan (1998) have shown that when this mechanism is removed and most or all meanings are potential referents in any given exchange, convergence is not attained. Thus, in these models the joint attention mechanism underpinning successful convergence has been implemented as a monolithic entity, with manipulation of this variable being limited to a dichotomous comparison of near perfect reliability to near zero reliability – full certainty of an intended referent versus high uncertainty.

Other studies have shown that, given certain learning strategies, convergence can also be attained in the absence of perfect context reduction. Models that simulate cross-situational learning and *obverter* signaling (Oliphant & Batali, 1997; Chapter 2, section 4.2.1) demonstrate that limiting context to a smaller number of meanings similarly enables conventionalization to take place (A.D.M. Smith, 2001; Vogt & Coumans, 2003). This reduction of context is presumed to result from various mechanisms for establishing joint attention (Smith et al., 2010). Thus, we can identify specific mechanisms capable of limiting context in this way and test how incorporating them into simulations affects convergence outcomes.

As described previously (Chapter 2, section 4.3), one aspect of establishing joint attention is the use and recognition of gesture. Nonarbitrary gestures carry meaning shared with or related to accompanying arbitrary words. Such gestures, while not supplying the full informational content of a spoken utterance, could serve to narrow the range of possible intended referents in a given context of communication by filtering out meanings that do not possess features included in a gestural signal. The potential contribution of gesture – in the form of related nonarbitrary representations that only partially indicate referents – in the emergence of shared vocabularies has not yet been modeled. It may be the case that partial context reduction like that provided by gesture-like communication has a substantial impact on establishing reference and constructing shared systems. In addition, the possible interaction of two modalities of signaling with different semiotic properties has not been modeled. If incorporating a second signaling channel that functions like gesture leads to successful convergence,

it would support the arguments made in this thesis regarding heterosemiotic multimodal communication and the conventionalization of novel symbols.

The following sections describe a series of simulations designed to investigate the dynamics of multimodal communication and the effects of a gesture-like component on emerging symbolic systems. The aim of this investigation is to understand the potential contribution of gesture and provide a more fine-grained analysis of the role of nonarbitrary representations in the conventionalization process. Thus, in regard to the interacting factors underlying grounding and convergence discussed in Chapter 1, section 4, we will attempt to minimize the operation of these other factors in order to isolate the effects of the gesture component as much as possible. The specifics of carrying this out will be described along with other details of model structure below.

Three models were created to address these issues. The first simulates communication through nonarbitrary representations alone, which will establish how this type of signaling can be implemented. The second is similar to previous models and simulates solely arbitrary signaling. The third combines the two modes of signaling, one arbitrary and one nonarbitrary, to simulate the heterosemiotic, multimodal nature of communication. The two individual models will show the particular potentials and limitations of each type of representation separately, and will also serve as a comparative tool to evaluate the dual model.

3. Model I – Nonarbitrary signaling

This model examines communication between identical agents that possess an inventory of signals that have nonarbitrary relationships to an inventory of meanings. All agents therefore already share all signals, as they are intrinsically and directly mapped to meanings by virtue of their forms. In this way, this model does not involve learning, evolving sign systems or emergence of any sort. Instead, it will show how differing levels of nonarbitrary information affect communicative accuracy.

3.1 Meaning and signal structure

Nonarbitrary representations are characterized by having a direct connection between their forms and the meanings they represent. Given that, we can implement nonarbitrary representations in this model by having meaning forms and signal forms be constituted by the same set of components. In this case, both meanings and signals will be binary vectors of varying lengths.

The size of the meaning inventory is determined by the number of bits that make up meanings. For example, designating meaning length at two bits produces a total of four distinct meanings: *00*, *01*, *10*, and *11*. A meaning length of three bits results in an inventory of eight meanings, and so on. We can mirror the partial nonarbitrariness of gesture by having the length of signals be shorter than the length of meanings (for present purposes we will designate that signals correspond to the end portion of meanings). A consequence of only partial signal-meaning coverage is that the same signal will overlap with more than one meaning; that is, no signal specifies a single individual meaning. We can also vary the amount of nonarbitrary information contained in a signal by varying the length of signals in proportion to the length of meanings. The length of a nonarbitrary signal determines how precisely it can specify a particular meaning; the longer the length, the less ambiguously a signal specifies a meaning. Tables 5.1-5.3 illustrate how signals are mapped to meanings using this method.

1.

Signals	Meanings
	000
0	*100*
	110
	010
	101
1	*001*
	011
	111

2.

Signals	Meanings
00	*000*
	100
10	*110*
	010
01	*101*
	001
11	*111*
	011

3.

Signals	Meanings
00	*0000*
	1000
	0100
	1100
10	*0010*
	0110
	1010
	1110
01	*0001*
	0101
	1001
	1101
11	*0011*
	0111
	1011
	1111

Tables 5.1-5.3 Examples of possible nonarbitrary communication systems; (1) one with three-bit meanings and one-bit signal, (2) one with three-bit meanings and two-bit signals and (3) one with four-bit meanings and two-bit signals. Signals represent the final bit or bits of meanings. Signals will never uniquely specify meanings, and the length of signals in proportion to meanings will determine the level of ambiguity at which signals can specify meanings.

We can vary certain parameters within this model to understand how the size of a communicative system and the amount of nonarbitrary information contained in signals affects communicative accuracy. These variables include meaning length (and thus inventory size), signal length (and thus inventory size and signal-meaning overlap), and context size. Given that agents are identical in their signal and meaning

inventories, communicative dynamics and accuracy will be constant for all population sizes. Therefore, population size will not be varied in this model, and simulations will consist of communication between two agents.

3.2 Communication dynamics

A communication event involves one agent sending a signal to communicate a given meaning to another agent. Communication takes place in a specified context of multiple meanings, the precise number of which can vary across simulations. Receiving agents must choose a meaning from this context as the sending agent's intended meaning, and this is done based on the information in the signal. Depending on the specified signal-to-meaning length ratio and the meanings comprising the context, some number of meanings may be removed as potential choices. Receiving agents then choose from the set of meanings (or meaning) that remains. If the meaning chosen matches the sending agent's intended meaning, the communication event is recorded as successful. Thus, the full process of a communication event involves the following steps:

1. A context of a certain size is created from a random selection of meanings from the full meaning inventory
2. A meaning to communicate (*MTC*) is chosen at random from the context
3. A 'speaker' sends a signal (*S*) that matches the corresponding portion of the MTC
4. A 'hearer' chooses a meaning (chosen meaning – *CM*) from the context
    a. Narrowing the context; all meanings that do not match *S* are removed
    b. If only one meaning remains, this is *CM*; If multiple meanings remain, one is chosen at random as *CM*
5. If *MTC* is the same as *CM*, the communication event is successful

Communicative accuracy (CA) is measured as the percentage of communication events that are successful in a series of exchanges. I will use 95% communicative accuracy as a standard for effective communication for this and subsequent models. The number of exchanges in a single simulation was set to 10,000 in order to create a large enough sample size reflect as closely as possible the underlying probabilities of overall communicative accuracy.

3.3 Results

Figures 5.1, 5.2 and 5.3 graph the communicative accuracy for simulations with meaning lengths of four, six and eight bits.

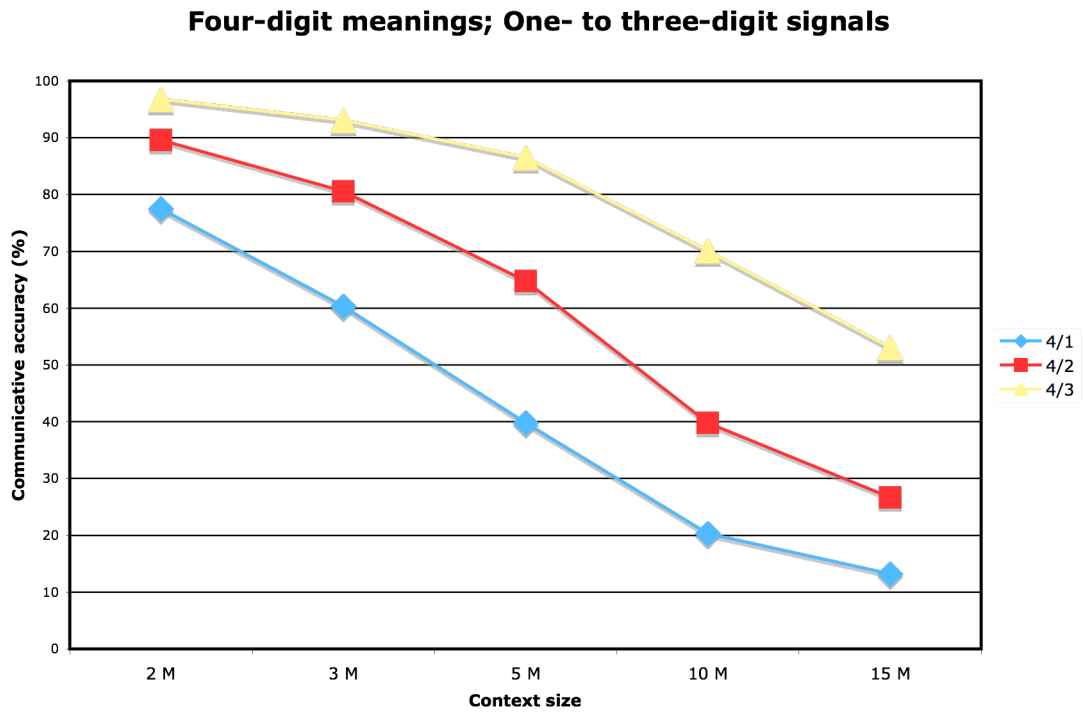


Figure 5.1 Communicative accuracy by context size for different levels of signal-meaning overlap. Each line represents a single simulation. 4/1 indicates that nonarbitrary signals overlap with meanings by one bit, 4/2 by two bits and so on.

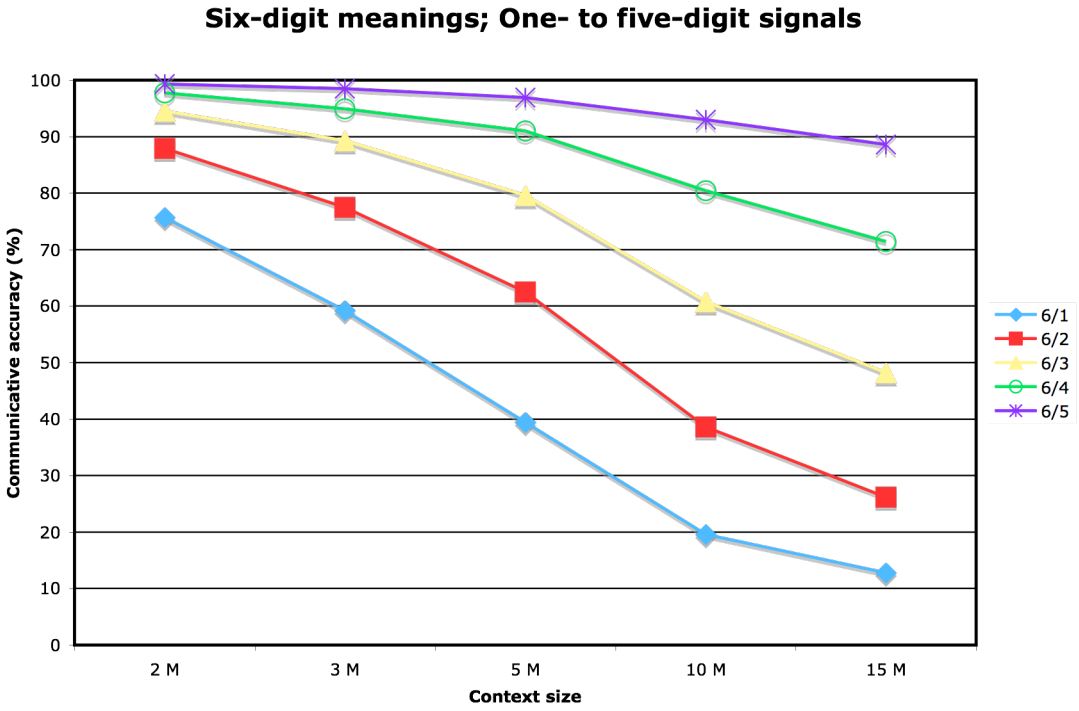


Figure 5.2 Communicative accuracy by context size for different levels of signal-meaning overlap. Each line represents a single simulation. 6/1 indicates that nonarbitrary signals overlap with meanings by one bit, 6/2 by two bits and so on.

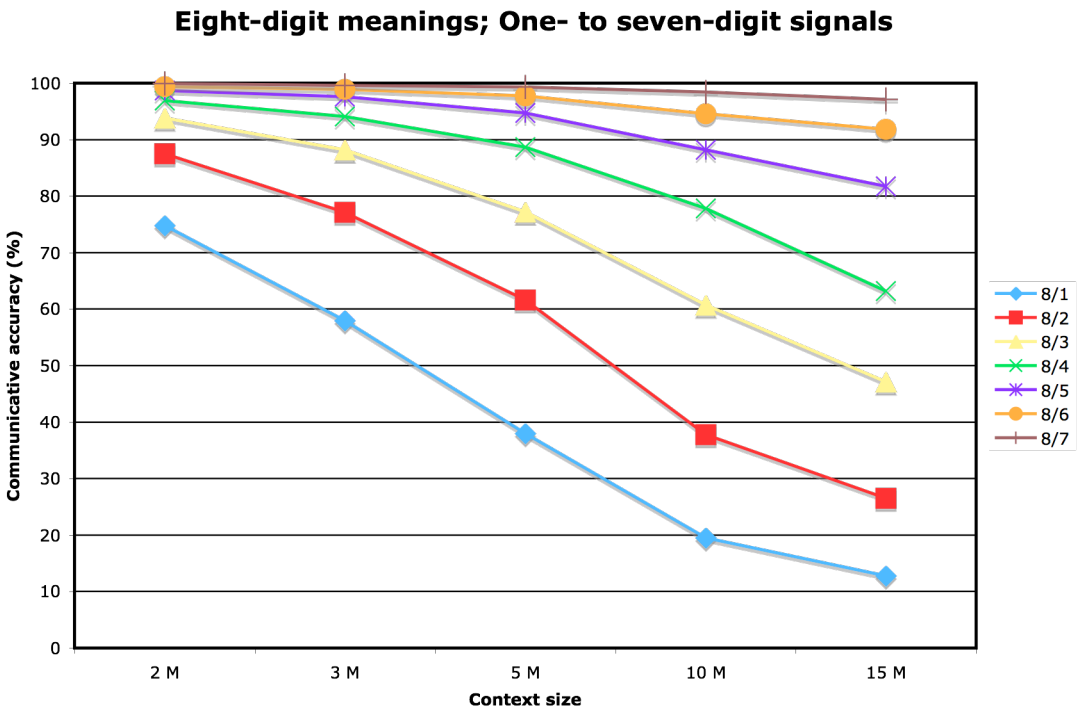


Figure 5.3 Communicative accuracy by context size for different levels of signal-meaning overlap. Each line represents a single simulation. 8/1 indicates that nonarbitrary signals overlap with meanings by one bit, 8/2 by two bits and so on.

For four-bit meanings, CA rises above 95% only when context consists of two meanings and with a signal overlap proportion of 75% - or when signals contain three of the four bits that constitute meanings. CA remains relatively high (93%) for the same meaning-signal ratio at a context size of three but does not reach our standard of effectiveness.

CA reaches 95% for six-bit meanings in contexts larger than two meanings. When the meaning-signal ratio is 6/5 – or 83% overlap – CA remains at effective levels up to a context size of five. A ratio of 6/4 reaches 95% in a context of three. A 50% overlap also meets the 95% standard, but only in contexts limited to two meanings.

For eight-bit meanings, CA reaches 95% in a wider range of conditions. Ratios from 8/4 to 8/7 are above this level when context size is set at two meanings. An overlap of 63% or greater maintains effective CA in contexts of five meanings. An 8/6 ratio reaches 95% in a context of ten, and an 8/7 ratio does so up to context of fifteen meanings.

The CA level reached for specific ratios changes as a function of meaning length. A 50% overlap does not meet the standard for four-bit meanings in a context of two, but the same ratio does for six- and eight-bit meanings. The longer meaning lengths also reach 95% in contexts with a greater number of meanings. Four-bit meanings only do so with a 75% overlap and context size set to two, while using six bits with the same ratio maintains effectiveness in contexts of three, and doing so for eight bits reaches 95% in contexts as large as ten meanings.

3.4 Discussion

These results demonstrate that partial information directly indicating meanings – like that embodied in nonarbitrary representations – can support effective communication. This finding is not surprising, as we would expect communication to be possible when interlocutors have access to each other's intended meanings, which is exactly what the nonarbitrary signals in this model aid in. Nevertheless, the efficacy of different levels of nonarbitrary information for meanings of different lengths in various context sizes informs how a model that incorporates these features can be structured.

The circumstances under which high communicative accuracy is maintained depend on context size, the length of meanings and the ratio of meaning length to signal length. As the number of bits comprising meanings increases, the potential for proportionally equivalent partial information in signals to support effective communication increases.

Meaning length reflects the number of distinct features meanings contain, and, in consequence, the total number of meanings. One way to think of this factor may be as i) the complexity of meanings and ii) the overall size of the meaning space over which communication operates. When this space is large and meanings are complex, effective communication is possible when signals contain half or a larger proportion of information indicating meanings. This holds for small context sizes, but also extends to much larger contexts when meanings are more complex and signals indicate a large amount of information. For example, an 88% overlap for eight-bit meanings maintains effective levels in context sizes up to 25 meanings (outside the range shown in Figure 3).

In relating these findings back to gesture, we can note that the relationship between signals and meanings in this model is much more straightforward than how gestures relate to their referents. That is, it is difficult – if not impossible – to quantify the amount of information contained in gesture and its proportion to the meanings they represent. The amount to which gestures represent features of their referents spans from the highly schematic to the highly detailed and precise depending on the type of gesture, the type of referent and the communicative demands involved. Nevertheless, we can reasonably assume that the number and complexity of meanings conveyed *through* gesture is high, given the rich and open-ended nature of human communication. In addition, gestures at least have the capacity to represent multiple and salient features of referents (Chapter 3, section 5). Based on these factors we can place gestural communication somewhere along the spectrum of communication systems modeled here. An appropriate comparison is probably to systems with a relatively higher number and complexity of meanings, and signals with moderate overlap with meanings. This characterization will be put to use later when carrying

out simulations on a model that combines nonarbitrary and arbitrary signaling in section 5 below.

Finally, effective communication here can be thought of as the capacity for information contained in signals themselves to enable agents to correctly identify intended meanings. These simulations show that this is possible even when signals do not fully represent all features of a referent. This context-limiting quality not only supports communication through nonarbitrary signals alone, but it could also serve a crucial function in conjunction with emerging arbitrary sign systems. We can better understand how accompanying nonarbitrary signals affect convergence by first designing a model of an independent arbitrary system and testing emergence outcomes in this system in isolation. The following section describes such a model.

4. Model II – Arbitrary signaling

This model will be similar to previous models of emergence discussed in detail in Chapter 2, section 4. Like those, this model will simulate the communicative and cultural dynamics of a population of agents that do not initially share signal-meaning mappings. Through repeated exchanges agents experience signal-meaning associations, learn them and thereby establish mappings. Whether or not the population of agents as a whole converges on the same mappings indicates if conventionalization has taken place.

Given that the ultimate aim of these simulations is to isolate the effects of an additional nonarbitrary signaling channel, certain characteristics of this model will differ from previous ones. First, the learning strategy that agents employ is simple associational learning – that is, there is no lateral inhibition or other in-built bias for one-to-one mappings (the effects of different learning strategies is discussed in depth by K. Smith, 2002). In addition, transmission is horizontal only within a closed group of agents. This design will enable us to understand the separate effects of nonarbitrary representations on emergence dynamics when we compare the outcomes of this and a combined model.

4.1 Meaning and signal structure

As this model will ultimately be combined with the previous nonarbitrary-based model, it will also use binary vectors as meanings. In this case, we want signals not to share any features with meanings. As such, signals will be strings of alphabetic letters. Even if signals contain multiple bits, however, they will not be analyzed by agents as containing discrete units. Agents will treat signals as holistic units regardless of the number or identity of the characters that comprise them.

The number of meanings and signals are both parameters of variation, so the exact inventory size for each and ratio between the two will vary depending on the simulation. Baronchelli, Dall'Asta, Barrat & Loreto (2006) have shown that a larger number of signals than meanings reduces the possibility for homonymy and aids convergence. Given that arbitrariness allows for many different forms to represent any given meaning, this design can include signal inventories that are large compared to meaning inventories, which will ensure failure to converge, if it does occur, is not the result of an in-built potential for homonymy.

4.2 Communication dynamics

All agents begin simulations with the same inventory of signals with which to send messages, but they do not initially have any mappings from signals to meanings – neither individually nor shared. As such, learning is a component of this model. A series of communication events will be simulated in which one agent attempts to communicate a certain meaning by choosing and sending a signal to another agent, who then attempts to understand the sending agent's intended meaning based on that signal. Agents have 'memories' in which meaning-signal pairs observed during communication events are stored with a frequency count. Each time a pair is encountered, the frequency count is increased. When one of a pair's component signal or meaning is encountered in subsequent exchanges, the probability that the complementary component of a stored pair is chosen during the communication process (either in signal selection for sending or meaning selection in receiving) is proportional to a pair's frequency count. That is, the higher a frequency count for a given pair, the more likely that the pair will be used to choose a signal or meaning

during a communication event that involves its component signal and meaning. In other words, pairs with the strongest association are not *guaranteed* to be chosen in any given exchange, but they have the greatest probability of being chosen, and this probability increases as frequency count increases. This method for meaning and signal selection was applied in order to introduce a small level of noise to the model. Noise of this sort is usually included in models of language processes and is intended to simulate typical aspects of natural communication, such as imperfect memory and/or other interference.

Unlike the nonarbitrary model above, agents in this case will not be identical and will have unique memories and vocabularies based on what signal and meanings they have experienced and learned while participating in communication events. As a result, it will be possible to vary the number of agents in a population to understand how this factor affects convergence.

Given the factors and dynamics described above, a communication event will involve the following steps:

1. A context is created from the full meaning inventory
2. A meaning to communicate (*MTC*) is chosen at random from the context
3. An agent is chosen at random from the population to be a speaker (*S*)
4. An agent is chosen at random to be a hearer (*H*; $H \neq S$)
5. *S* sends a signal to convey the *MTC*
    a. *S* checks stored knowledge for meaning-signal pair(s) that contain *MTC*
        i. If none are found, *S* chooses a signal from inventory at random
        ii. If some are found, a signal is selected using the method described above.
6. *H* chooses a meaning from the context
    a. Checks stored knowledge for a meaning-signal pair that contains the signal and any meanings from the context
        i. If none found, *H* chooses a meaning at random from context
        ii. If some are found, a meaning is chosen (*CM*) using the method described above.
7. *S* and *H* update their stored knowledge
    a. *S*: if (signal, *MTC*) in stored knowledge, increase frequency count by 1

H: if (signal, *CM*) in stored knowledge, increase frequency count by 1

b. *S*: if (signal, *MTC*) not in stored knowledge, add pair and assign frequency count of 1
   H: if (signal, *CM*) not in stored knowledge, add pair and assign frequency count of 1

Note that a consequence of this process is that a given exchange can result in the speaker and hearer storing different meaning-signal pairs. Because there is no error feedback, hearers will store their chosen meaning with the observed signal regardless of whether it is the correct meaning. Likewise, speakers do not have access to hearers' choices and will store the pair they initially selected to send regardless of the outcome of the communication event. This simulation will demonstrate if shared mappings can arise under these dynamics and under what conditions that occurs.

Through a series of exchanges as described in steps 1-7, agents build vocabularies, and these stored vocabularies can then be used to test communicative accuracy. After a simulation of communication and learning is completed, the following process is repeated.

1. Steps 1-6 from above
2. If *MTC* = *CM* the communication event is successful

Repeating these steps simulates many communication events between agents that have formed vocabularies and in the process keeps track of how many are successful. Communicative accuracy is calculated as the proportion of successful communication events to total communication events. Using our standard of 95% accuracy, a result at this level or higher indicates that agents can communicate successfully because their vocabularies overlap to a large degree. In other words, this measure of communicative accuracy will indicate the extent to which agents share meaning-signal mappings.

With the model's structure and dynamics laid out, we can now identify parameters of variation. These include meaning inventory size, signal inventory size, context size, population size and number of exchanges.

4.3 Results

A number of different simulations were run varying meaning inventory size and signal inventory size along with other variations. Similar trends were found for different meaning-signal inventory ratios, but we will focus on simulations that attained the highest communicative accuracy levels. Figure 5.4 shows the results of a simulation with a meaning inventory of 50 and a signal inventory of 264. These results reflect populations that have engaged in 100,000 exchanges, which was the point at which general and lasting patterns were found to stabilize across individual simulations.

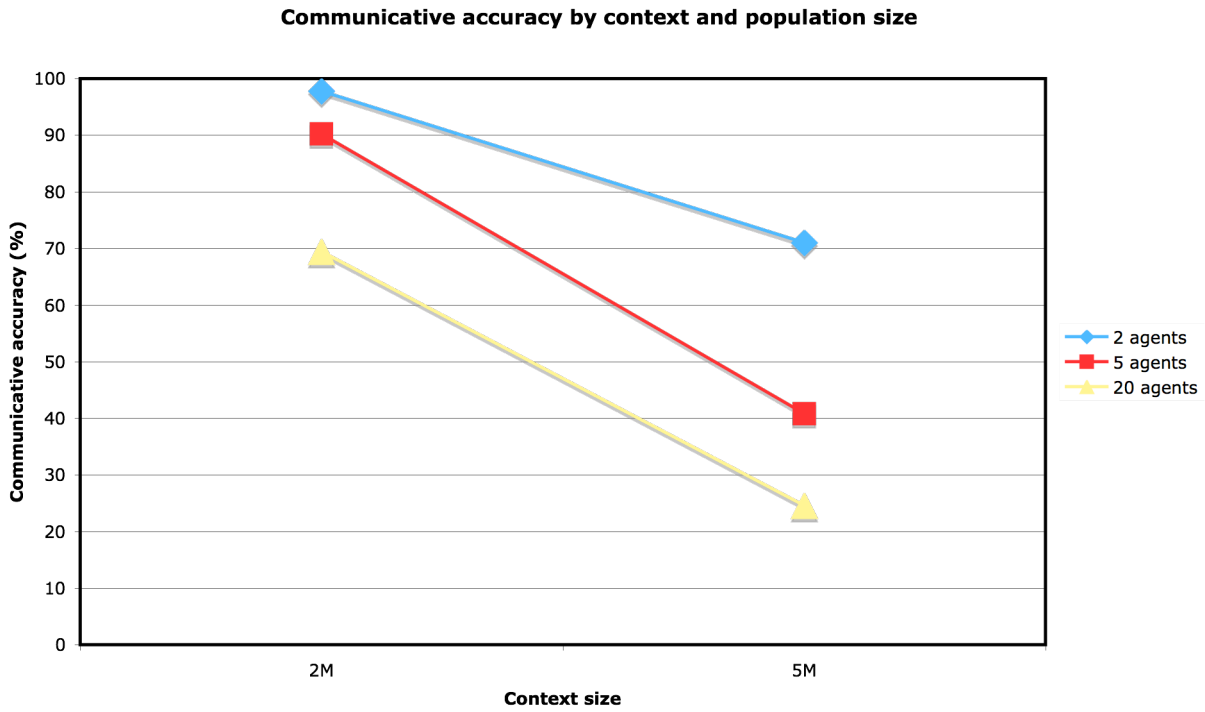


Figure 5.4 Communicative accuracy by context size for an arbitrary-only signaling system with 50 meanings and 264 signals. Each color represents a different populations size.

Communicative accuracy reaches 95% only when population size and context size are at the lowest level simulated: two meanings and two agents, respectively. No population larger than two agents converged on a shared vocabulary. The effect of population size changes as context size increased. The difference between accuracy levels for population sizes of two and five become more pronounced, rising from 8%

in contexts of two meanings to 25% in contexts of five meanings. In contrast, the difference in accuracy levels between population sizes of five and twenty narrows, dropping from a difference of 31% to 21%.

4.4 Discussion

These results demonstrate that the learning strategy and communicative dynamics implemented in this model on their own only support the emergence of effective communication, and thereby convergence, in the most limited of circumstances. We can compare the results to Vogt & Coumans (2003) study, which also tested convergence outcomes for context sizes greater than one and different population sizes. Their model implemented cross-situational learning and the obverter signaling strategy of the type described in Chapter 2 section 4, which are more powerful mechanisms for establishing mappings between meanings and signals than the simple associational learning and signal selection strategies employed here. Simulations of this model showed that 95% communicative accuracy can be reached when contexts contain five meanings for populations of two and six agents, and populations of ten agents near this level (approximately 90%). In comparison, communicative accuracy in the current model with the same context size only rose to 75% for a population of two agents, and further dropped to 44% with a population of five.

Present results are also in line with what we would expect based on K. Smith's (2002) study. He showed that, even when context is limited to a single meaning paired with a single signal, shared systems do not emerge in populations of agents that learn using a simple associational strategy like that applied here. These simulations demonstrate that similar results hold when generational turnover and transfer are not involved, though in this case convergence does occur in the very constrained conditions of context and population sizes of two.

In relating these results to naturalistic communication, it is clear that a population size of two individuals communicating in contexts of only two meanings is not particularly realistic. Even when context size remains relatively small final accuracy levels decrease substantially. Without a grounding mechanism to reduce context sizes or other strategies to cope with ambiguity, convergence is not a likely outcome. Overall,

then, the attributes represented in this model do not on their own constitute a robust or viable model for convergence. These results can, however, serve as a baseline for comparison to a model that incorporates a grounding mechanism based on nonarbitrary representations. The following section will describe a design that combines the two types of communication simulated in this and the preceding model.

5. Model III – Multi-channel heterosemiotic signaling

This model will combine the structure of the two preceding models to determine how partial information like that provided by nonarbitrary representations affects grounding and convergence. As in Model II, learning and signal/meaning selection based on arbitrary mappings are factors and will be implemented using the same methods described above (section 4.3).

5.1 Meaning and signal structure

Like both preceding models, meanings will be binary vectors. Agents will have two inventories of signals with which to construct messages. One will be a fixed and shared inventory of nonarbitrary signals like that described in section 3.2. The other will be a set of arbitrary signals like that described in section 4.2 that are initially unassociated with any meanings. In any given exchange, the agent tasked with sending a message will choose a signal of each type and send both together.

5.2 Communication dynamics

Given that both types of signals are included in this model, signal and meaning selection combines the two processes above. A sending agent will choose the nonarbitrary signal that most closely represents a meaning and send this along with an arbitrary signal chosen based on that agent's past communication experience (when available). In turn, the agent receiving this composite signal can use the nonarbitrary portion to identify the intended meaning or reduce the context by eliminating meanings that do not match it. In addition, the receiving agent can use the arbitrary portion of the signal to further inform their choice based on their own communicative history.

Thus, a given communication event proceeds as follows:

1. A context is created from the full meaning inventory
2. A meaning to communicate (*MTC*) is chosen at random from the context
3. An agent is chosen at random from the population to be a speaker (*S*)
4. An agent is chosen at random to be a hearer (*H*; $H \neq S$)
5. *S* sends a composite signal to convey the *MTC*
    a. Chooses a nonarbitrary signal (Sg_1) that matches the corresponding portion of the *MTC*
    b. Chooses an arbitrary signal (Sg_2)
        i. Checks stored knowledge for a meaning-(arbitrary)signal pair that contains *MTC*
            1. If no, chooses an arbitrary signal from inventory at random
            2. If yes, selects arbitrary signal based on method described above (section 4.2)
6. *H* chooses a meaning (*CM*) from the context
    a. Narrowing the context: all meanings that do not match Sg_1 are removed; if only one meaning remains, this is *CM* and step 6b is skipped.
    b. Chooses meaning from remaining context using Sg_2
        i. Checks stored knowledge for a meaning-signal pair that contains Sg_2 and any meanings from the context
            1. If no, chooses a meaning at random
            2. If yes, selects a chosen meaning (*CM*) based on method described above (section 4.2)
7. *S* and *H* update their stored knowledge
    a. *S*: if (Sg_2, *MTC*) in stored knowledge, increase frequency count by 1
    H: if (Sg_2, *CM*) in stored knowledge, increase frequency count by 1
    b. *S*: if (Sg_2, *MTC*) not in stored knowledge, add pair and assign frequency count of 1
    H: if (Sg_2, *CM*) not in stored knowledge, add pair and assign frequency count of 1

Through a series of exchanges and learning, agents build arbitrary signal vocabularies. The extent to which these vocabularies are shared and convergence has taken place can be evaluated by repeating the following process:

1. Steps 1-6 from the *arbitrary-only* model above (section 4.2)[18]
2. If *MTC* = *CM* the communication event is successful

[18] Note that nonarbitrary signals are not included in communication events. Their exclusion makes it possible to measure independently the communicative accuracy and convergence status of agents' constructed arbitrary signaling systems.

As in the arbitrary-only model, we will consider a communicative success rate of 95% as indicating that agents have converged on shared signal-meaning mappings.

As in the arbitrary-only model, agents will not be identical and therefore population size is again a parameter to vary in simulations, along with context size, meaning inventory size, arbitrary signal inventory size, and nonarbitrary signal-meaning overlap and inventory size. Given the considerations discussed in section 4.4 regarding ratio of arbitrary signals to meanings, this section will focus on simulations with a proportionally large arbitrary signal inventory. In addition, given the considerations regarding information content of nonarbitrary representations, it will focus on simulations with six-bit meanings, which will allow a fine-grained analysis of different overlap levels.

5.3 Results

Figures 5.5 and 5.6 show the communicative accuracy levels attained after 50,000 exchanges for simulations with 64 six-bit meanings, 264 arbitrary signals and varying nonarbitrary signal lengths between four and five bits. These are shown alongside the results for the arbitrary-only model in section 4 (represented in black). For context sizes of two, communicative accuracy reaches 95% for a number of population sizes and signal-meaning overlap ratios. All multi-channel simulations attain this level with populations of five, and those with an overlap of 67% and 83% do as well in populations of twenty agents. With context size set to five, simulations with overlap ratios of 33% or higher reach 95% in populations of two agents. For populations of five agents, those with a ratio of 50% or higher continue to reach this level. For populations of twenty none reached 95% accuracy, though a simulation with 83% overlap was only slightly below at 94.4%.

Another way to look at these results is in how many of the six circumstances possible by varying context size and population size convergence was reached. Convergence occurred in the multi-channel model in two out of six with a 17% overlap, three out of six with a 33% overlap, four out of six with a 50% overlap and five out of six with both a 67% and 83% overlap. In comparison, communicative accuracy levels indicate convergence was reached in the arbitrary-only model for only one in six simulations.

The vocabularies of a sample population under similar conditions to the one examined in section 4.3 above also differs from the arbitrary model. For a population of two agents and context size of five, 99% of the resulting vocabularies are shared, compared with 87% in the arbitrary model. If we look at words within vocabularies, on average between the two agents 63% of the 64 meanings have synonymous signals. In contrast to the arbitrary model, in which a single meaning could have as many as eleven synonyms, the highest number of signals mapped to a meaning in this model was three. Finally, if we look at the individual signals, 31% are homonyms and map to multiple meanings. In comparison, in the arbitrary model 41% of signals were homonymous.

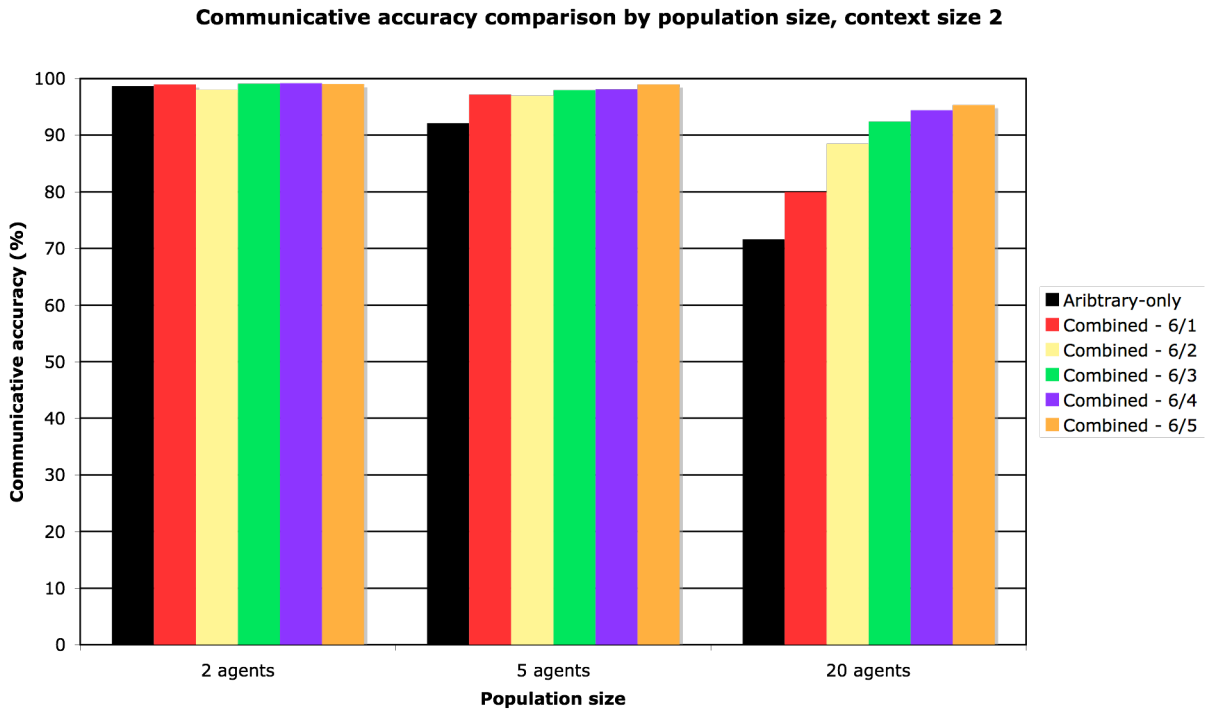


Figure 5.5 Multi-channel model results for context size of two and six-bit meaning length with varying population sizes. Colored bars represent different nonarbitrary signal overlap ratios from 17-83%. Arbitrary-only results are shown in black.

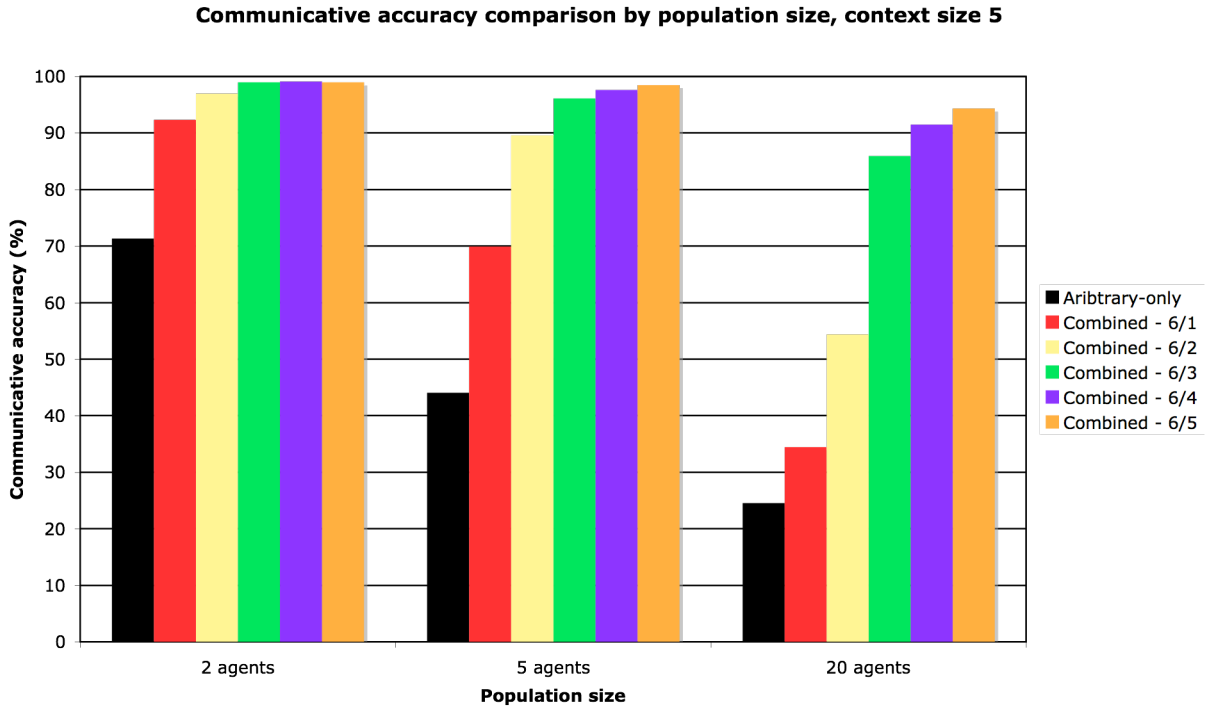


Figure 5.6 Multi-channel model results for context size of five and six-bit meaning length with varying population sizes. Colored bars represent different nonarbitrary signal overlap ratios from 17-83%. Arbitrary-only results are shown in black.

5.4 Discussion

These results demonstrate that incorporating an additional channel of signaling that provides information directly related to meanings greatly increases the ability for populations to converge on shared arbitrary signal-meaning mappings. This occurs as a result of the context reducing effect of nonarbitrary signals, which in turn enables successful grounding for the accompanying arbitrary signals.

Convergence was reached in these simulations in the absence of a bias for one-to-one mappings (K. Smith, 2002; Steels & Kaplan, 1998; Vogt & Coumans, 2003) or cross-situational learning strategies (A.D.M. Smith, 2001; Vogt & Coumans, 2003) included in previous studies. Equipped with only a simple learning strategy of association, which on its own cannot reliably instigate convergence (section 4 above; K. Smith, 2002), populations of agents eventually converge on shared vocabularies of arbitrary signal-meaning mappings. Thus, we can reason it is the action of the accompanying nonarbitrary signals that is responsible for this effect.

The vocabularies themselves demonstrate why communicative accuracy is higher than the arbitrary model alone. The agents in Model III do in fact converge on a nearly fully shared vocabulary (99% overlap), while those in Model II do not (87%). There are also much lower levels of homonymy and synonymy resulting from Model III simulations. Moreover, agents share the same homonyms and synonyms when they do exist, whereas many of these are not shared by agents from Model II simulations. Although Model III vocabularies are not optimal communication systems because of these ambiguous signals, they are nevertheless effective, as evidenced by communicative accuracy rates. We can assume that the reason effective communication is still possible for these populations and not those in Model II is because i) agents share the same vocabularies and ii) those vocabularies have comparatively much fewer ambiguous signals.

The breakdown of differing levels of nonarbitrary information content in signals reveals that this device plays a more complex role than has been shown in the zero versus maximum ambiguity conditions modeled previously (Steels & Kaplan, 1998). These results show that convergence can still take place when the quantity of information in nonarbitrary signals is low to moderate – as low as a 17% overlap in some cases (see Figure 5). Moreover, even when low informational content in nonarbitrary signals does not result in full convergence here, it nevertheless increases communicative accuracy well above the arbitrary-only base model. This particularly holds for larger context sizes and population sizes, as illustrated in Figure 6. If we consider the combined effects of more sophisticated and additional learning strategies, like those implemented in previous models and shown to be employed naturally by human learners in behavioural experiments (Smith et al., 2010), the capacity for nonarbitrary signaling to enable convergence becomes even greater. We can reason that convergence outcomes would be resilient across a broader range of context sizes and population sizes, and more closely reflecting circumstances of real-world communication.

One point to address is that these results may appear surprising in light of K. Smith's (2002) finding that agents using the learning strategy modeled here did not converge on a shared vocabulary or attain high communicative accuracy. However, certain

differences in population structure and signal-meaning structure can likely account for the successful outcome of this model. First, transmission in K. Smith's (2002) simulations was vertical, with agents receiving a single exposure to a previous agent's signal and meaning pairs, then producing their acquired signal and meaning pairs a single time to the next agent. In the present model, the population of agents remains stagnant and transmission is horizontal, which means agents send and receive messages repeatedly over time. It may be that this cumulative process allows shared signal-meaning pairs, when they do arise, to be reinforced and strengthened in a way that cannot occur through vertical transmission like that modeled by K. Smith. Another difference in the models is in the ratio of signal inventory to meaning inventory. In K. Smith's model, the number of signals is equal to the number of meanings, while in the present model the number of signals is greater than meanings (by a factor of four). As described above, a larger ratio of signals to meanings decreases the potential for homonymy, which in turns aids the convergence process (Baronchelli et al., 2006). Thus, the greater ratio modeled here may also be in part responsible for the observed higher levels of communicative accuracy, and together with population structure can account for the difference in results with K. Smith's (2002) model.

These results provide additional support for the idea that symbolic communication can be built upon nonarbitrary representations. The schematized and partial version of nonarbitrary representation applied here resembles gestural communication of the kind described in Chapter 3 that typically accompanies speech. This model therefore reflects two fundamental aspects of human communication: multimodality and heterosemiotic representation distributed between the modalities. Incorporating these features into simulations in itself enabled convergence in a range of circumstances and constitutes an additional ecologically and evolutionarily viable mechanism for models of convergence. Present results therefore demonstrates how a natural feature of human communication – nonarbitrary representations via gesture – can contribute to the emergence of shared symbolic systems and, together with past research, provides a robust model and explanation for conventionalization through cultural processes alone.

6. General discussion

Overall the results of Models I-III support the claim made in Chapter Three that nonarbitrary gestures could have provided sufficient grounding for arbitrary vocalizations. Various levels of partial information made available in accompanying nonarbitrary representations was shown to be capable of reducing uncertainty to a point at which conventionalization of arbitrary sign-meaning mappings could take place. Moreover, this conventionalization occurred through the operation of unintentional cultural and communicative processes. Thus, these findings support the broader arguments of this thesis that multimodal communication and nonarbitrary representation are powerful explanatory resources for understanding how symbolic communication emerged in human evolution.

One assumption of these arguments is that gesture provides the kind and quantity of information represented in the nonarbitrary signals of Models I and III. One avenue for future related research would be to verify that natural gestures do in fact enable people to correctly identify referents under similar circumstances. Other recent research has shown that computational models and their results can be successfully recreated in the laboratory (Kirby, Cornish & Smith, 2008). Following these efforts, it would be possible to similarly design experiments to reflect the dynamics of Model III and test what kind of gestures people spontaneously produce, their grounding potential and their capacity to underpin conventionalization of arbitrary signs.

Finally, these simulations indicate that there may be new and additional approaches to exploring the emergence of shared vocabularies in computational modeling by taking into account the particular features of embodied human communication and their consequences. The potential for multiple channels of signaling, different semiotic roles of channels and the temporal and semantic relationship between the two open up a variety of possible conditions and factors that could influence convergence. Investigating these factors will enrich our understanding of conventionalization processes and the emergent properties of complex systems generally as well as the cultural ramifications of natural, embodied human communication specifically.

7. Conclusion

The simulations presented in this chapter extend findings from previous studies that show populations can converge on shared arbitrary signals when those signals are sufficiently grounded. I have shown that incorporating an additional channel of signaling that reflects the properties of gestural representation into traditional models of symbolic communication enables the emergence of shared vocabularies under a range of conditions. The presence of nonarbitrary signals allows the accompanying arbitrary signals to be mapped to the same meanings by agents across populations, forming a shared symbolic system that is communicatively functional in its own right. These results support the claim made in Chapter 3 of this thesis that the nonarbitrary gestures produced in multimodal communication could serve as an in-built grounding mechanism for accompanying vocalizations, and together with other cognitive and transmission factors known to aid convergence this model provides a robust avenue for conventionalization.

CHAPTER 6
Conclusion

1. Scope of focus and research covered

This thesis set out to investigate the emergence of symbolic communication in human evolution. Given that this question is one rooted in human physical and social experience, an embodied perspective was taken in order to frame the issue accurately. This approach removes the question from the abstract and situates it in terms of the particular causally relevant features of lived human experience.

In addition to taking an embodied perspective, the question is one that involves communicative signs and evolutionary processes, and therefore required the incorporation of ideas and principles from semiotics and evolutionary theory. An examination of sign theory indicated that i) interpretation of signs as arbitrarily related to their meanings and ii) the conventionalization of those signs between individuals are the factors underlying symbolic communication. As such, an explanation of emergence must account for both of these factors. A plausible evolutionary scenario for emergence must also adhere to general evolutionary principles, notably those of parsimony and continuity. It was also shown that one should avoid appealing to deliberate invention or purported communicative advantages as spurring the development of symbols.

With these ideas in mind, Chapter 2 reviewed and analyzed research on emerging sign systems. Evidence from experimental sign studies and emerging sign languages show that communication is established through nonarbitrary signs. Furthermore, results indicate that symbolic reinterpretations of these early signs do not arise via intentional creation by individuals, but instead nonarbitrary signs become symbolic when they are transmitted to others who did not participate in their creation and development. These results thus further support the argument that an explanation for the emergence should first explore the potential for 'blind', unintentional forces to instigate the interpretation of signs as symbolic before implicating the insight of users or other

deliberate actions to construct a conventional, culturally-based method of communication.[19]

Emerging sign systems research also spoke to the issue of conventionalization. A review of computational models simulating the construction of shared symbolic vocabularies across populations showed that mechanisms must be present to establish and maintain joint attention. Joint attention is what provides interlocutors access to each other's intended meanings when no information is available in a sign's form and thereby enables the grounding of arbitrary symbols. Thus, it was determined that an explanation for the conventionalization component of the emergence of symbols must account for a mechanism or mechanism by which novel symbols could be grounded in this way and adopted across a population.

Chapter 3 applied the general processes identified in Chapter 2 to the specific case of symbolic communication in human evolution. Human communication takes place predominantly through the vocal and gestural modalities, and these modalities have distinct semiotic capacities. While both in theory are equal in capacity for arbitrary representation, gesture has a greater capacity for nonarbitrary representation than vocalization. Gestures can be used to richly and more precisely depict a wider variety of referents than vocalizations, and these asymmetrical semiotic capacities are mirrored in the complementary functions the modalities serve in typical communication. The vocal modality is the dominant channel for symbolic communication, while gestures are primarily used for nonarbitrary representation. Reviewed research also showed that vocal symbolic communication does not occur in isolation but is instead tightly temporally and semantically coupled with simultaneously produced nonarbitrary gestures. Because symbols are psychologically and communicatively intertwined with bodily signs in this way, the question of emergence was reframed as one of how the specific multimodal, heterosemiotic speech-gesture system observed today arose. In addition, the particular qualities of vocal versus gestural representation have implications for interpretation and

[19] This is not to imply that communicators' actions would not be deliberate in the sense that individuals intentionally attempt to be understood in any given act of communication. It is on the cultural, population level at which the feasibility of deliberate actions becomes untenable – that is, communicators deliberately coordinating to create a conventional, symbolic vocabulary.

conventionalization processes, and the explanatory potential of these qualities was taken into consideration in the remainder of the chapter that addressed the question of emergence directly.

2. A theory of emergence

Existing theories of emergence were reviewed and critiqued according to the ideas presented in this thesis. Taken together, these proposals were shown not to address sufficiently the question of emergence in terms of the critical factors of interpretation and conventionalization, either by omission or implicit appeals to evolutionarily and/or psychologically unviable mechanisms. I then argued that theoretical gaps and problematic scenarios could be resolved by mining the explanatory potential of embodied communication and sign interpretation processes. First, the predominance of symbolic vocalizations would be a natural consequence of interpretation processes, as the vocal modality's lower capacity for nonarbitrary representation relative to that of gesture would translate into more vocalizations being interpreted as arbitrary and symbolic. Second, the reinterpretation and subsequent conventionalization of novel symbols could take place in normal communicative and transmission processes. These new arbitrary vocalizations would have an organic, in-built grounding mechanism available in the simultaneously produced semantically related nonarbitrary gestures. The presence of gesture would thus enable interlocutors to identify correctly each other's intended messages and thereby create the opportunity for the accompanying vocalizations to be interpreted correctly and adopted as shared symbols. This scenario represents a more comprehensive and theoretically satisfactory explanation as it i) takes into account a more accurate characterization of embodied human communication and ii) does not require additional pressures for symbols or the innovation and insight of individuals in order for a coupled symbolic vocal and nonarbitrary gesture system like that observed in modern behavior to emerge. In this way, it maximizes the explanatory potential of taking an embodiment perspective and is more closely in line with evolutionary principles.

3. Empirical contribution

These claims rest on certain assumptions regarding i) how the vocal and gestural components of multimodal utterances are interpreted and processed, and ii) the ability for information like that made available in gesture to adequately ground arbitrary signs. Chapter 4 described experiments undertaken to test whether multimodal messages similar to the novel arbitrary-vocal plus nonarbitrary-gestural combinations used in the purported context of emergence are processed as a psychologically unified whole as speech and gesture utterances are in language today. Results indicate that such unified processing still occurs under these conditions, with participants attending to and learning vocalizations that are largely uninformative relative to co-occurring gestures. These findings suggest that novel arbitrary vocalizations would be attended to, learned and adopted through typical communicative processes without the need for additional pressures favoring vocal symbols. Chapter 5 described a series of computational simulations designed to model heterosemiotic multimodal communication and test the capacity for nonarbitrary signals of varying informational content to ground paired arbitrary signals, and thereby enable populations to converge on shared vocabularies through cultural processes alone. Results showed that nonarbitrary signals in isolation – that is, without the additional aid of more sophisticated learning and signaling strategies – can sufficiently reduce uncertainty and allow agents to correctly identify each other's intended meanings, which in turn allows arbitrary signals to conventionalize across a population of agents. These findings suggest that nonarbitrary gestures, together with other known mechanisms that support grounding, represent a robust model for the conventionalization of symbols. This model is constituted by factors inherent to the communicative process itself and would be in operation during normal transmission dynamics. As such, it avoids implicating additional and/or intention-based forces in order to explain how shared arbitrary signs could arise and therefore represents an evolutionarily and psychological plausible explanation.

4. Theoretical contribution

By explaining how symbolic communication could emerge out of 'blind' interpretation and transmission processes inherent to typical communicative

dynamics, the arguments made in this thesis remove the need to identify additional pressures and conditions in human evolution that would have spurred the development of symbols. That is, given the preconditions of bodily mimesis and related capacities, it is possible for conventional vocal symbols to arise through natural communicative and cultural processes alone. This line of reasoning highlights the capacity for bodily mimesis as a critical juncture in human evolution, one that both represents a major biological and cultural change from precedents and also has substantial and far-reaching ramifications for the potential subsequent development and elaboration of resulting communicative systems. Doing so allows us to focus our attention and research efforts on identifying the causes behind this seminal event in human history. In addition, the position argued for in this thesis provides a foundation for other research areas that take as a starting point conventional symbolic communication.

5. Future research

Finally, the central ideas of this thesis point to a number of new avenues of research. Areas that can be explored include: more precisely measuring the capacity for nonarbitrary gestures to convey messages of varying complexity and types of meaning; experimentally testing in the lab which transmission dynamics spur a transition from nonarbitrary to arbitrary signs (assuming the analysis of previous experiments presented in Chapter 2); experimentally testing in the lab the processes and mechanisms by which novel arbitrary signs become conventionalized; testing whether modifications of the experimental conditions tested in Chapter 4 produce results like those for speech and gesture demonstrating beneficial memory effects for multimodal over unimodal utterances; and, if these beneficial effects hold, exploring whether and how they could affect conventionalization processes.

Appendix – Publications

Brown, J.E. (2010). Coordinated multi-modal expression and embodied meaning in the emergence of symbolic communication. In *The Evolution of Language: Proceedings of the 8th International Conference (EVOLANG)*, (eds) A.D.M. Smith, M. Schouwstra, B. de Boer & K. Smith (pp. 375-376). World Scientific.

COORDINATED MULTI-MODAL EXPRESSION AND EMBODIED MEANING IN THE EMERGENCE OF SYMBOLIC COMMUNICATION

J. ERIN BROWN

Language Evolution and Computation, PPLS, The University of Edinburgh, 3 Charles Street, Edinburgh EH8, UK

The study of language evolution has a long tradition of connecting gestures to language origins (Condillac, 1746; Hewes, 1973). Modern theories point to gesture as the solution to a central problem: the emergence of symbolic communication. Prominent versions (Arbib, 2005; Corballis, 2002; Tomasello, 2008) share three critical features: i) early forms of communication consisted of pointing and pantomiming; ii) these gestures then became conventionalised and arbitrary, or symbolic; iii) at some point, the symbolic channel 'switched' and vocalisations became the dominant channel for symbolic communication. I agree that i) is a plausible stage in language evolution but contend that points ii) and iii) are less likely, as they do not follow the evolutionary principles of parsimony and continuity, nor do they provide a satisfactory explanation for the relationship between speech and gesture as it exists today (McNeill, 2005). In addition, arguments for this scenario rely on questionable assumptions regarding early hominid gestural and vocal abilities, the vocal channel's greater potential for creating arbitrary symbols and the role of speech in the instruction of manufacturing techniques.

Although these accounts recognise the powerful representational potential of gesture and consider the advantages of an additional, distinct modality of communication, they do not appear to fully appreciate the synergistic potential of both modalities together nor the limitations of a single modality on its own. If mimetic gestures became symbolic as postulated, the power of their 'natural' meaning would have been lost. Moreover, distributing meaning expression between symbolic and nonarbitrary forms provides cognitive and communicative benefits in language production and comprehension (Goldin-Meadow et al., 2001; Kelly et al., 1999), an advantage that would be sacrificed if gestures transitioned into arbitrary symbols. Though it is not claimed nonarbitrary gestures disappeared during this transition, this scenario does not allow for the same simultaneous nonarbitrary-and-arbitrary signaling distributed across modalities that would enable the cognitively demanding task of forming symbols.

Another problem for these theories is the 'switch' to vocalisations as the dominant vehicle for symbolic communication. If a symbolic gestural system arose, it would have been hugely advantageous and caused evolutionary forces to move toward manual signed language, thus making it very unlikely for speech to evolve (Emmorey,

2005). In addition, an evolutionary scenario in which signaling types shift between modalities entails multiple and significant evolutionary transitions.

A careful consideration of gesture research and the nonarbitrary nature of human communication can contribute substantially to our understanding of language origins. The representational power of gesture alone is not sufficient to explain how arbitrary forms came to carry meaning, as claimed in current gestural origins theories. It is the *coordinated multimodality* of human expression that provides the opportunity for bodily manifestations of meaning to be transferred to co-occurring vocal signals. If nonarbitrary gestures co-occurred with vocalisations early in hominid history, it presents an opportunity for sounds to become symbolic while preserving gesture's 'natural' meaning and retaining the cognitive and communicative benefits of gesturing. In this view, symbols arose in the modality in which they still occur today, thus obviating a 'switch' in symbolic channel in the course of human evolution.

References

Arbib, M., 2005. From monkey-like action recognition to human language: An evolutionary framework for neuroscience. *BBS, 28*, 105-124.

Condillac, E.B., 1746. Essai sur l'origine des connassances humaines, ouvrage ou l'on reduit a un seul principe tout ce concerne l'entendement. In *Oeuvres philosophiques de Condillac*. Paris: Georges LeRoy.

Corballis, M., 2002. *From hand to mouth: The origins of language*. Princeton University Press.

Emmorey, K., 2005. Sign languages are problematic for a gestural origins theory of language evolution. *BBS, 28*, 130-131.

Hewes, G.W., 1973. Primate communication and the gestural origin of language. *Current Anthropology, 12*, 5-24.

Kelly, S.D., Barr, D.J., Breckinridge Church, R. & Lynch, K., 1999. Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and Memory. *Journal of Memory and Language, 40*, 577-592.

McNeill, D., 2005. *Gesture and Thought*. Cambridge University Press.

Tomasello, M., 2008. *Origins of Human Communication*. MIT Press.

References

Ahlner, F. & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies, 38*, 298-348.

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: we think, therefore we gesture. *Language and Cognitive Processes, 15*, 593-613.

Alibali, M.W., Heath, D.C. & Myers, H.J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169-188.

Arbib, M. A. (2003) The Evolving Mirror System: A Neural Basis for Language Readiness. In S. Kirby & M.H. Christiansen (eds.), *Language Evolution* (pp. 182-200). Oxford University Press.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioural and Brain Sciences , 28*, 105-167.

Arbib, M.A., Liebal, K. & Pika, S. (2008). Primate vocalization, gesture and the evolution of human language. *Current Anthropology, 49*, 6, 1053-63.

Armstrong, D. F., Stokoe, W. C., and Wilcox, S. (1995). *Gesture and the Nature of Language.* Cambridge University Press.

Aziz-Zadeh, L., Wilson, S.M., Rizzolatti, G. & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology, 16*, 1818-1823.

Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., and Steels, L. (2006) Sharp Transition towards Shared Vocabularies in Multi-Agent Systems. *Journal of Statistical Mechanics*, P06014.

Batali, J. (1998) Computational simulations of the emergence of grammar. In Hurford, J. R. and Studdert-Kennedy, M. and Knight C. (eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases*, pp. 405-426. Cambridge University Press.

Bates, E. and Dick, F. (2002), Language, gesture, and the developing brain. *Developmental Psychobiology, 40*, 293–310.

Bavelas, J. B., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language 58*, 495-520.

Beattie, G., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica, 123*, 1-30.

Boulenger, V., Roy, A., Paulignan, Y. Deprez, V., Jeannerod, M. & Nazir, T.A. (2006). Cross-talk between language processes and overt motor behavior in the first 200 msec of processing. *Journal of Cognitive Neuroscience, 18*, 10, 1607-1615.

Boyd, R. & P. J. Richerson. (1985). *Culture and the Evolutionary Process*. University of Chicago Press.

Burling, R. (2005). *The Talking Ape: How Language Evolved*. Oxford University Press.

Cangelosi, A. & Parisi, D. (2002). *Simulating the Evolution of Language*. London: Springer.

Capirci, O. & Volterra, V. (2008). Gesture and speech: The emergence and development of a strong and changing partnership. *Gesture, 8*, 1, 22-44.

Cartmill, E. A. & Byrne, R. W. (2010). Do ape gestures have specific meanings?: Shifting the focus from flexibility to semanticity. In *The Evolution of Language: Proceedings of the 8th International Conference (EVOLANG)*, (eds) A.D.M. Smith, M. Schouwstra, B. de Boer & K. Smith (pp. 377-378). World Scientific.

Cavalli-Sfornza, L. and M. Feldman. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press.

Chawla, P. & Krauss, R.M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology, 30*, 580-601.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again.* MIT Press.

Clark, A. (2008). *Supersizing the mind: Embodiment, action and cognitive extension.* Oxford University Press.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, & I. Sag (eds.), *Elements of discourse understanding*, pp. 10-63. Cambridge University Press.

Clark, H.H., 1996. *Using Language*. Cambridge University Press.

Cole, J., Gallagher, S. & McNeill, D. (2002). Gesture following deafferentation: a phenomenologically informed experimental study. *Phenomenology and the Cognitive Sciences. 1*, 49-67.

Corballis, M. C. (2002). *From hand to mouth: The origins of language*. Princeton University Press.

Corbalis, M. C. (2003). From Hand to Mouth: The Gestural Origins of Language. In S. Kirby & M.H. Christiansen (eds.) *Language Evolution*, (pp. 201-218). Oxford University Press.

Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain & Language 112*, 25-35.

Cornejo, C., Simonetti, A.I., Aldunate, N., Ceric, F., Lopez, V. & Nunez, R.E. (2009). Gesture and metaphor comprehension: Electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain and Cognition, 70*, 42-52.

Cuskley, C. and Kirby, S. (Forthcoming). Synaesthesia, cross-modality and language evolution. In: Simner, J. and Hubbard E.M. (eds.). *Oxford Handbook of Synaesthesia*.

Danesi, M., (1999). *Of Cigarettes, High Heels and Other Interesting Things: An Introduction to Semiotics*. Palgrave Macmillan.

de Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics, 28*, 441-465.

Deacon, T.W., 1997. The Symbolic Species: The Co-evolution of Language and the Brain. W.W. Norton & Company.

Deacon, T. (2009). Relaxed Selection and the Role of Epigenesis in the Evolution of Language. *Oxford Handbook of Developmental Behavioral Neuroscience* (pp. 730–752). Oxford University Press.

Deb Roy. (2005). Semiotic Schemas: A Framework for Grounding Language in the Action and Perception. *Artificial Intelligence, 167*, 1-2, 170-205.

DeLoache, J.S. (1995). Early Understanding and Use of Symbols: The Model Model. *Current Directions in Psychological Science, 4*, 4, 109-113.

Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard University Press.

Donald, M. (1998) Mimesis and the executive suite: Missing links in language evolution. In J.R. Hurford, M. Studdert-Kennedy and C. Knight (eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge University Press.

Eco, U., (1979). *A Theory of Semiotics*. Indiana University Press.

Efron, D. (1972). *Gesture, race, and culture*. Mouton.

Ekman, P., Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*, 124-129.

Emorrey, K. (2005). Sign languages are problematic for a gestural origins theory of language evolution. Commentary to Arbib (2005). *Behavioural and Brain Sciences, 28:2, 130-131*.

Fay, N. & Lim, S., 2010. From hand to mouth: An experimental simulation of language origin. In *The Evolution of Language: Proceedings of the 8th International*

Conference (EVOLANG), (eds) A.D.M. Smith, M. Schouwstra, B. de Boer & K. Smith. World Scientific.

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

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

Fex, B. & Mansson, A.C. (1998). The use of gestures as a compensatory strategy in adults with acquired aphasia compared to children with specific language impairment (SLI). *Journal of Neurolinguistics, 11*, 191-206.

Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology, 18*, 2, 185-205.

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

Fodor, J.A. (1975). *The Language of Thought*. Harvard University Press.

Fontana, S. (2008). Mouth actions as gesture in sign language. *Gesture, 8*, 1, 104-123.

Frishberg, N., (1975). Arbitrariness and iconicity: Historical change in American Sign Language. *Language, 51*, 3, 696-719.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. Cogn. Sci. 29, 737–767.

Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford University Press.

Garrod, S., and Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science, 1*, 292–304.

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

Garrod, S., Fay, N., Rogers, S., Walker, B., and Swoboda, N. (2010). Can iterated learning explain the emergence of graphical symbols? *Interaction Studies 11*, 33–50.

Gasser, M., (2004). The origins of arbitrariness in language. *Annual conference of the Cognitive Science Society, 26.*

Gentilucci, M., Corballis, M.C. (2006). From manual gesture to speech: a gradual transition. *Biobehavioral Review, 30*, 949-960

Gil, D. (2006). How much grammar does it take to sail a boat? (Or, what can material artefacts tell us about the evolution of language?). In Smith, A. D. M., Smith, K.,

Ferrer-i-Cancho, R. (eds.), *The evolution of language: Proceedings of the 7th international conference.* World Scientific Press.

Glenberg, A. M. & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 3, 558-565 .

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

Goldin-Meadow, S. (2003a). *The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us About How All Children Learn Language.* Psychology Press.

Goldin-Meadow, S. (2003b). *Hearing Gesture: How Our Hands Help Us Think.* Harvard University Press.

Goldin-Meadow, S., So, W.-C., Ozyürek, A., & Mylander, C. The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences, 105, 27*, 9163-9168.

Gould, S. J. (1977). The return of hopeful monsters. *Natural History, 86*, 24, 30.

Gould, S.J. (1989). *Wonderful Life: The Burgess Shale and the Nature of History.* W. W. Norton.

Gould, S.J. (1996). *Life's Grandeur.* Trafalgar Square.

Gould, S. J. & Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London, B, 205*, 1161, 581-98.

Hadar, U. (1986) Forcefield analogy for communications involving movement of the head. *Semiotica, 62*, 279-296.

Hadar, U. & Pinchas-Zamir, L. (2004). The semantic specificity of gesture: Implications for gesture classification and function. *Journal of Language and Social Psychology, 23*, 2, 204-214.

Hanlon, R.E., Brown, J.W. & Gerstman, L.J. (1990). Enhancement of naming in nonfluent aphasia through gesture. *Brain and Language, 38*, 298-314.

Hauk, O., Johnsrude, I. & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*, 301-307.

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

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

Heine, B. & Kuteva, T. (2008). *The Genesis of Grammar: A Reconstruction.* Oxford University Press.

Hewes GW. (1973). Primate communication and the gestural origins of language. *Current Anthropology, 14*, 5-24

Hinton, L., Nichols, J. & Ohala, J.J. (eds.) (1994). *Sound Symbolism*. Cambridge University Press.

Hockett, C. F. (1960). Logical considerations in the study of animal communication. In *Animal Sounds and Communications*, (eds.) W. Lanyon and W. Tavolga. (Washington: American Institute of Biological Sciences), pp. 392–430.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience, 19*, 1175.

Holler, J. & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language & Cognitive Processes, 24*, 267-289.

Hopkins, W. D., & Savage-Rumbaugh, E. S. (1991). Vocal communication as a function of differential rearing in Pan paniscus. Some preliminary findings. *International Journal of Primatology, 12*, 559-583.

Hopkins W.D., Taglialatela, J.P. & Leavens, D.A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of humans. *Animal Behaviour, 73*, 281-286.

Horner, V & Whiten, A (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal Cognition, 8*, 164-181.

Hostetter, A. B., Alibali, W. M., & Kita, S. (2007). I see it in my hand's eye: Representational gestures are sensitive to conceptual demands. *Language and Cognitive Processes, 22*, 3, 313-336.

Hubbard, A.L., Wilson, S.M., Callan, D.E. & Dapretto, M. (2009). Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping, 30*, 3, 1028-1037.

Husserl, E. (1931). *Ideas: General Introduction to a Pure Phenomenology.* Trans. W.R. Boyce Gibson. London: Allen and Unwin.

Husserl, E. (1960). *Cartesian Meditations: An Introduction to Phenomenology.* Trans. Dorian Cairns. The Hague: Martinus Nijhoff.

Hutchins, E. and Johnson, C. M. (2009). Modeling the Emergence of Language as an Embodied Collective Cognitive Activity. *Topics in Cognitive Science, 1*, 523--546.

Iverson, J. M., & Goldin-Meadow, S. (2001). The resilience of gesture in talk: Gesture in blind speakers and listeners. *Developmental Science, 4*, 416-422.

Iverson, J.M. & Thelen, E. (1999). Hand, mouth, and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies, 6*, 19-40.

Keller, R. A, (1998). *Theory of Linguistic Signs*. Oxford University Press.

Kelly, S. D., Barr, D., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory & Language, 40*, 4 577-592.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*, 253-260.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes, 24*, 313-334.

Kelly, S. D., Ozyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*, 260-267.

Kemmerer, D., Chandrasekaran, B. & Tranel, D. (2007). *Cognitive Neuropsychology, 24*, 1, 70-114.

Kendon, A. (2004). *Gesture: Visible Action as Utterance.* Cambridge University Press.

Kirby, S. (1999). *Function, Selection and Innateness: the Emergence of Language Universals*. Oxford University Press.

Kirby, S., Dowman, M. and Griffiths, T. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences, 104*, 12, 5241-5245.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Science U.S.A, 105*, 10681–10686.

Kita, S. (2008). World-view of protolanguage speakers as inferred from semantics of sound symbolic words: A case of Japanese mimetics. In N. Masataka (ed.), *Origins of Language* (pp.25-38). Springer.

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes, 24*, 2, 145-167.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*, 16-32.

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes, 22*, 8, 1212-1236.

Kita, S., & Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes, 24*, 5, 761-775.

Kita, S, Kantarzis, K, & Imai, M. (2010) Children Learn Symbolic Words Better: Evolutionary Vestige of Sound Symbolic Protolanguage pp. In (eds.) A.D.M. Smith, M. Schouwstra, B. de Boer & K. Smith *The Evolution of Language: Proceedings of the 8th International Conference (EVOLANG*) (pp. 206-213). World Scientific.

Kita, S., Ozyurek, A., Allen, S. & Ishizuka, T (2010) Early Links Between Iconic Gestures and Sound Symbolic Words: Evidence for Multimodal Prototolanguage. In (eds.) A.D.M. Smith, M. Schouwstra, B. de Boer & K. Smith *The Evolution of Language: Proceedings of the 8th International Conference (EVOLANG*) (pp. 429-430). World Scientific.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In M. Zanna (ed.), *Advances in experimental social psychology* (pp. 389-450). Academic Press.

Kwek, J. & Kendon, A. (1991). Occasions for sign use in an Australian aboriginal community. Sign Language Studies 20, 71, 143–160

Lott, P. (1999). *Gesture and aphasia*. Peter Lang.

Lust, B. & Foley, C., (2004). *First Language Acquisition. The Essential Readings.* Wiley-Blackwell.

Maricchiolo, F., Gnisci, A., Bonaiuto, M. & Ficca, G. (2009). Effects of different types of hand gestures in persuasive speech on receivers' evaluations. *Language & Cognitive Processes, 24*, 2, 239-266.

Marler, P., Evans, C.E. & Hauser, M.D. (1992). Animal vocal signals: reference, motivation or both? In H. Papoušek & U. Jürgens (eds.) *Nonverbal Vocal Communication: Comparative and Developmental Approaches* (pp. 66-86). Cambridge University Press.

Masson, M. E. J., Bub, D. N., & Newton-Taylor, M. (2008). Language-based access to gestural components of conceptual knowledge. *Quarterly Journal of Experimental Psychology, 71*, 869-882.

Mayberry, R. I. & Nicholadis, E. (2000). Gesture reflects language development: Evidence from bilingual children. *Current Directions in Psychological Science, 9*, 192-196.

Mayberry, R. I., & Jaques, J. (2000). Gesture production during stuttered speech: Insights into the nature of gesture-speech integration. In D. McNeill (ed.), *Language and Gesture* (pp. 199-213). Cambridge University Press.

McNeill, D. (1992). Hand and mind: What gestures reveal about thought. University of Chicago Press.

McNeill, D. (2000) (ed.). *Language and Gesture*. Cambridge University Press.

McNeill, D. (2005). *Gesture and Thought.* University of Chicago Press.

McNeill, D., Cassell, J. & McCullough, K.E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language & Social Interaction, 27*, 3, 223-237.

McNeill, D., Bertenthal, B., Cole, J., & Gallagher, S. (2005). Gesture-first, but no gestures? Commentary to Arbib (2005). *Behavioural and Brain Sciences , 28*, 2, 138-139.

McNeill, D., Duncan, S. D., Cole, J., Gallagher, S., and Bertenthal, B. (2008) Growth points from the very beginning. Interaction Studies, 9:1, 117-132.

Meguerditchian, A., & Vauclair, J. (2010). Investigation of gestural vs vocal origins of language in nonhuman primates: distinguishing comprehension and production of signals. *The Evolution of Language: Proceedings of the 8th International Conference (EVOLANG*), (eds) A.D.M. Smith, M. Schouwstra, B. de Boer & K. Smith. World Scientific.

Meir, I., W. Sandler, C. Padden, & M. Aronoff (2010). Emerging sign languages. In M. Marschark and P. Spencer, (eds.). *Oxford Handbook of Deaf Studies, Language, and Education. 2*. Oxford University Press.

Meir, Irit (2002a). A cross-modality perspective on verb agreement. *Natural Language and Linguistic Theory, 20*, 2, 413-450.

Melinger, A. & Levelt, W.J.M (2005). Gesture and the communicative intention of the speaker. *Gesture 4*, 2 119-141.

Merleau-Ponty, M. (1962). *Phenomenology of Perception*. Trans. Colin Smith. London: Routledge and Kegan Paul.

Merleau-Ponty. M. (1963). *The Structure of Behavior*. Trans. Alden Fisher. Boston: Beacon Press.

Mesoudi, A. & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B, 363*, 3489-3501.

Mesoudi, A., Whiten, A. & Laland, K.N. (2006) Towards a unified science of cultural evolution. *Behavioral and Brain Sciences 29*, 4, 329-383.

Mithen, S.J. (2004). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. Weidenfeld & Nicolson.

Morgan, L.H. (1877). *Ancient Society.* MacMillan & Company.

Morsella, E., & Krauss, R.M. (2004). The role of gestures in spatial working memory and speech. *American Journal of Psychology, 117*, 3, 411-424.

Namy, L. L. (2008). Recognition of Iconicity Doesn't Come for Free. *Developmental Science, 11*, 841-846.

Nicoladis, E., Pika, S., & Marentette, P. (2010). Are number gestures easier than number words for preschoolers? *Cognitive Development, 25*, 247-261.

Noë, A. & O'Regan, J.K. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*, 5, 883-917.

Noë, A. (2004). *Action in Perception*. MIT Press.

Nuckholls, J.B. (1999). The case for sound symbolism. *Annual Review of Anthropology, 28*, 225-252.

Oliphant, M. and Batali, J. (1997). Learning and the emergence of coordinated communication. *The newsletter of the Center for Research in Language*, 11, 1.

Olson, D.R. (1996). Towards a psychology of literacy: on the relations between speech and writing. *Cognition, 60*, 83-104.

Ong, W., 1982. *Orality and Literacy: The Technologizing of the Word.* Methuen.

Ozyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *Journal of Cognitive Neuroscience, 19*, 605-616.

Paget, R. (1930). *Human Speech: Some Observations, Experiments, and Conclusions as to the Nature, Origin, Purpose, and Possible Improvement of Human Speech.* Routledge.

Parrill, F., Bullen, J., & Hoburg, H. (2010). Effects of input modality on speech-gesture integration. *Journal of Pragmatics 42*, 11, 3130-3137

Peirce, C. S. (1931–1958). Collected papers of Charles Sanders Peirce. Vols. 1–8. C. Hathorne, P. Weiss, & A. Burks (eds.). Harvard University Press.

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

Pine, K. J. Bird, H. & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science, 10*, 6, 747-754.

Pizzuto, E.A. & Capobianco, M. (2007). The link and differences between deixis and symbols in children's early gestural-vocal system. In K. Liebal, C. Müller & S. Pika (eds.), *Gestural Communication in Nonhuman and Human Primates* (pp. 161-179). John Benjamins.

Pulvermüller, F., Shtyrov, Y. & Ilmoniemi, R. (2005). Brain signature of meaning access in action word recognition. *Journal of Cognitive Neuroscience, 17*, 1-9.

Quine, W. (1960). *Word and object.* MIT Press.

Ramachandran, V.S. & S. Blakeslee (1998). *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. William Morrow

Rauscher, F. B., Krauss, R. M., & Chen, Y. (1996). Gesture, speech and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*, 226-231.

Richerson, P. J. & Boyd, R. (2005). *Not By Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press.

Rime, B., Schiaratura, L., Hupet, M. & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation & Emotion, 8*, 311-325.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21*,188-194.

Rueschemeyer, S.A., Lindemann, O., van Rooij, D., van Dam, W. & Bekkering, H. (2010). Effects of intentional motor actions on embodied language processing. *Experimental Psychology, 57*, 4, 260-266.

Sandler, W., Meir, I., Padden, C., and Aronoff, M. (2005). The emergence of grammar: systematic structure in a new language. *Proceedings of the National Academy of Science U.S.A. 102*, 2661–2665.

Sato M., Mengarelli, M., Riggio, L., & Gallese V. (2008). Task related modulation of the motor system during language processing. *Brain and Language, 105*, 83-90

Saussure, F. (1916). Nature of the Linguistic Sign. In C. Bally & A. Sechehaye (eds) Cours de linguistique générale. McGraw Hill Education.

Savage-Rumbaugh, S., Shanker, S., & Taylor, T. (1988). *Apes, Language and the Human Mind*. Oxford University Press.

Schmandt-Besserate, D. (1996). How Writing Came About. University of Texas Press.

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

Sebeok, T.A. (1994). *Signs: An Introduction to Semiotics*. Toronto University Press.

Senghas, A. (1995). Children's contribution to the birth of Nicaraguan Sign Language. Ph. D. dissertation, Massachusetts Institute of Technology. Distributed by MIT Working Papers in Linguistics.

Senghas, A. (2005). Language emergence: Clues from a new Bedouin sign language. *Current Biology, 15*, 12, 463-465.

Simner, J., Cuskley, C. and Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception, 39*, 4, 553-569.

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

Skipper, J., Goldin-Meadow, S., Nusbaum, H., & Small, S. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language, 101*, 260-277.

Slocombe, K. E. and Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of the Sciences, 104*, 43, 17228-17233

Slocombe, KE, Waller BM, Liebal K. (2011). The language void: the need for multimodality in primate communication research. *Animal Behaviour, 81*, 919-924.

Smith, A.D.M. (2001). Establishing communication systems without explicit meaning transmission. In J. Kelemen and P. Sosík (eds.) Advances in Artificial Life (6th ECAL Conference). 381-390.

Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science, 14*, 1, 65-84.

Smith, K., Smith, A.D. M., Blythe, R. A., & Vogt, P. (2006). Cross-situational learning: a mathematical approach. In P. Vogt, Y. Sugita, E. Tuci, & C. Nehaniv (eds.) *Symbol Grounding and Beyond: Third International Workshop on the Emergence and Evolution of Linguistic Communications* (pp. 31-44). Springer-Verlag.

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011) Cross-situational learning: an experimental study of word-learning mechanisms. *Cognitive Science, 35,* 480-498.

Smith, L.B. & Gasser, M. (2005). The development of embodied cognition: Six lessons from babies. *Artificial Life, 11*, 13-30.

Sonesson, G. (1997). The ecological foundations of iconicity. In: Rauch, Irmengard; Carr, Gerald (eds.), *Semiotics Around the World: Synthesis in Diversity. Proceedings of the Fifth International Congress of the IASS* (pp. 739–742). Mouton de Gruyter.

Sperber, D. & Wilson, D. (1987). Precis of relevance: Communication and cognition. *Behavioral and Brain Sciences, 10*, 697-754.

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

Steels, L. & Kaplan, F. (1998) Stochasticity as a source of innovation in language games. In C. Adami and R. Belew and H. Kitano and C. Taylor, (eds.), *Artificial Life VI*. MIT Press.

Steels, L. & Vogt, P. (1997). Grounding adaptive language games in robotic agents. In C. Husbands and I. Harvey (Eds.), *Proceedings of the Fourth European Conference on Artificial Life*. Cambridge Ma. and London: The MIT Press.

Steels, L., Kaplan, F., McIntyre, A., and Van Looveren, J. (2002) Crucial Factors in the Origins of Word-Meaning. In A. Wray (ed.), The Transition to Language (pp. 252-271). Oxford University Press.

Stefanini, S., Bello, A., Caselli, M.C., Iverson, J.M., & Volterra, V. (2009). Co-speech gestures in a naming task: Developmental data. *Language and Cognitive Processes, 24*, 168-189.

Sterelny, K. & Griffiths, P.E. (1999). *Sex and Death: An Introduction to Philosophy of Biology*. University of Chicago Press.

Taglialatela J.P., Savage-Rumbaugh, S. & Baker, L.A. (2003). Vocal production by a language-competent bonobo, Pan paniscus. *International Journal of Primatology, 24*, 1, 1-17.

Tanner, J. E. (2004). Gestural phrases and gestural exchanges by a pair of zoo-living lowland gorillas. *Gesture, 4*, 1, 1-24.

Tanner, J. E., & Byrne, R. W. (1996). Representation of action through iconic gesture in a captive lowland gorilla. *Current Anthropology, 37*, 1, 162-173.

Tettamanti, M., Buccino, G., Saccuman, M.C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S.F. & Perani, D. (2005). Listening to action-related sentences actives fronto-parietal motor circuits. *Journal of Cognitive Neuroscience, 17*, 273-281.

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

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind.* Harvard University Press.

Thompson, R.K.R., Oden, D.L., & Boysen, S.T. (1997). Language-naïve chimpanzees (Pan troglodytes) judge between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal behavior processes, 23*, 31-43.

Tomasello, M., 1999. *The Cultural Origins of Cognition*. Harvard University Press.

Tomasello, M., 2008. Origins of Human Communication. MIT Press.

Tomasello, M., & Farrar, J. (1986). Joint attention and early language. *Child Development, 57*, 1454–1463.

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. MIT Press.

Vogt, P. (2000a). Bootstrapping grounded symbols by minimal autonomous robots. *Evolution of Communication, 4(1)*, 89-118.

Vogt, P. (2000b). *Lexicon Grounding on Mobile Robots*. Unpublished Ph.D. thesis, Vrije Universiteit Brussel.

VOGT, P. (2001). The impact of non-verbal communication on lexicon formation. In B. Kröse,M. De Rijke,G. Schreiber and M. Van Someren (Eds.), *Proceedings of the 13th Belgian/Netherlands Artificial Intelligence Conference, BNAIC'01.*

Vogt, P., & Coumans, H. (2003). Investigating social interaction strategies for bootstrapping lexicon development. *Journal of Artificial Societies and Social Simulation, 6*, 1.

Wagner, S., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language, 50*, 395-407.

Wesp, R., Hesse, J., & Keutmann, D. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology, 114,* 591–600.

Willems, R.M., Hagoort, P. & Casasanto, D. (2010). Body-specific representations of action verbs: Neural evidence from right- and left-handers. *Psychological Science, 21*, 1, 67-74.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *12* (4): 625-636

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

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP Study. *Brain and Language, 101*, 234-245.

Wundt, W.M. (1916). *Elements of Folk Psychology*. G. Allen and Unwin Ltd.

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., and Braun, A. R. (2009) Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences, 106*, 49, 20664-20669.

Yamaguchi, C. & Izumi, A. (2008). Vocal learning in nonhuman primates: Importance of vocal contexts. In N. Masataka (ed.), *Origins of Language*. Springer.

Zenon W. & Pylyshyn, Z.W. (1980). Cognitive representation and the process-architecture distinction. *Behavioral and Brain Sciences, 3*, 154-169

Zlatev, J. (2007). Embodiment, language and mimesis. In T. Ziemke, J. Zlatev & R. Franck (eds.), *Body, Language, Mind. Vol 1: Embodiment*, pp. 297-337. Mouton de Gruyter.

Zlatev, J. (2008). From proto-mimesis to language: Evidence from primatology and social neuroscience. *Journal of Physiology, Paris, 102*, 1-3, 137-151.

Zlatev, J., Persson, T., & Gärdenfors, P. (2005). Bodily Mimesis as the "Missing Link" in Human Cognitive Evolution, LUCS 121. Lund: Lund University Cognitive Studies.