INFANTS' UNDERSTANDING OF SIGNS: LINKING SOUNDS AND CONCEPTS

by

HO HENNY YEUNG

B.S., Duke University, 2003

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA

August 2006

© Ho Henny Yeung, 2006
Abstract

Spoken words are semiotic signs: speech sounds which function to signify concepts in the world. Infants' emerging understanding of the semiotic relation between sounds and concepts is explored here. First, basic questions from the philosophy of semiotics are recast as broader questions relevant for the psychological study of word-learning in infancy. Developmental aspects of one particular semiotic principle are explored in an empirical study which investigates whether infants bi-directionally link sounds and concepts. Previous research has suggested that this link is present, at least in one direction, very early in infancy: the presence of a word influences the structure of concepts, even before infants are able to robustly learn the referents of new words. The converse question is the topic of the present work: does the presence of an object when hearing speech sounds facilitate discrimination and identification of these sounds as potential word forms? Two studies provide evidence that for 9-month-old infants this is indeed the case. In Study 1, infants were exposed to a familiarization phase in which sounds from two non-native phonetic categories were contrastive, paired concordantly with two objects. In Study 2, another group was exposed to a familiarization phase in which the sounds where not contrastive, paired discordantly with the objects. Infants only succeeded in discriminating the contrast in the former case, suggesting that they must have used the link between sound and object to categorize the phonetic information. At an age when infants begin to have difficulty discriminating non-native phonetic contrasts, this research shows that infants can re-learn to discriminate this contrast if given evidence that it might signify two different words. This provides one example of how semiotic principles can be applied to research in language acquisition.
# Table of Contents

Abstract ......................................................................................................................... ii

Table of Contents ........................................................................................................ iii

List of Tables .................................................................................................................... iv

List of Figures .................................................................................................................. v

List of Illustrations .......................................................................................................... vi

1. Some philosophy and psychology of signs ................................................................. 1
   1.1 Introduction .............................................................................................................. 1
   1.2 Semiotic Principles .................................................................................................. 3
      1.2.1 Saussure ........................................................................................................... 4
      1.2.2 Peirce ............................................................................................................... 5
      1.2.3 Principles ......................................................................................................... 6
   1.3 Semiotic acquisition ............................................................................................... 8
      1.3.1 Signs ............................................................................................................... 8
      1.3.2 Arbitrariness ................................................................................................... 16
      1.3.3 Systems .......................................................................................................... 24
   1.4 Concluding Summary ............................................................................................ 27

2. Empirical work on signs ............................................................................................ 29
   2.1 Introduction .............................................................................................................. 29
      2.1.1 Word-learning at 12 months and beyond ......................................................... 31
      2.1.2 Words-as-cues-for-concepts ........................................................................... 36
      2.1.3 Concepts-as-cues-for-words .......................................................................... 40
      2.1.4 Early Speech Perception .................................................................................. 43
   2.2 Study 1 .................................................................................................................... 46
      2.2.1 Method ............................................................................................................. 47
      2.2.2 Results .............................................................................................................. 54
      2.2.3 Discussion ....................................................................................................... 55
   2.3 Study 2 .................................................................................................................... 56
      2.3.1 Method ............................................................................................................. 57
      2.3.2 Results .............................................................................................................. 59
      2.3.3 Discussion ....................................................................................................... 61
   2.4 General Discussion ............................................................................................... 64

Notes ............................................................................................................................... 69

Bibliography ................................................................................................................... 70

Appendix A ...................................................................................................................... 79
List of Tables

Table 1: Conceptual summary for psychological applications of semiotic theory. .......... 28

Table 2: Conceptual summary of studies in early word-learning .................................. 31
List of Figures

Figure 1: Format contours plotted for each individual token of /da/ .................. 49
Figure 2: Results from Study 1 ........................................................................ 55
Figure 3: Results from Study 2 ........................................................................ 60
Figure 4: Summary of Studies 1 & 2 .................................................................. 61
List of Illustrations

Illustration 1: One possible pairing of objects and sounds ........................................ 51
Illustration 2: Sample of Pairings used in Study 2 .................................................. 57
1. Some philosophy and psychology of signs

1.1 Introduction

A long-standing question of interest in the field of developmental psychology is, "how do infants and young children learn words?" In reviewing existing research on this topic, a series of nested and more specific research questions arise, many of which have generated a cohesive research program independently of themselves. Restricting ourselves to the study of infants below 15 months of age, word-learning research has tended to ask three types of questions:

- How do infants learn to perceive and identify the forms that words can have? (Jusczyk, 1997; Saffran, Werker, & Werner, 2006)
- How do infants learn to perceive and identify the things which words can label? (Waxman, 2004; Xu, 2005)
- How do infants actually learn to link word forms with things? (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Golinkoff, Shuff-Bailey, Olguin, & Ruan, 1995; Werker, Cohen, Lloyd, Casasola, & Stager, 1998)

Given these three types of questions, one reasonable question might be: is the way that psychological research partitions the task of word-learning indicative of the task which actually faces an infant? In other words, and without trivializing the difficulty of each of these tasks, does "word-learning" mean that an infant learns word forms, the categories which word forms can label, and then must work out which word form corresponds with which conceptual category? Indeed, this view essentially echoes
positions taken by nominalist philosophers, who argued that possible names and real
objects were constrained by ontology—that learning a word depends on the actual
existence of a real-world object (i.e., a cat) or concept (i.e., the tangible feeling of anger;
the practice of democracy, etc.) and a physical utterance (e.g., [kʰət]). This is exactly the
sort of question which is the topic of this thesis; what exactly does “learning a word”
entail? How might other philosophical traditions inform us vis-à-vis the use and
understanding of words?

Words, of course, are a subset of what philosophers and linguistics have often
called “signs,” or more generally, any physical form that represents or stands for an
object, idea, thing, etc.—the study of which has been termed, “semiotics” (Sebeok, 1994).
Not coincidentally, of course, the three kinds of research questions mentioned above
correspond to three epistemological distinctions that philosophers and semioticians make
when describing the structure of signs: the physical or perceived form of a sign itself (i.e.,
the signifier or the representamen), that which it signifies (i.e., the signified or the
object), and the psychological relation which holds between the former two (i.e., the
signification or the interpretant). These terms come concurrently, but also independently,
from both Ferdinand de Saussure (1916) and Charles S. Peirce (1908), two founders of
modern semiotics1.

The claim that psychological research in word-learning has tended to cluster
around epistemological distinctions made in the philosophy of signs seems
uncontroversial. Yet, beyond this tripartite division, it seems that developmental research,
particularly in the field of word-learning, has tended to avoid deeper philosophical
inquiry into what principles characterize the use of signs. Stated differently, while
psychological approaches have broken down the study of words in the same way that a semiotician might examine a sign, developmental research in word-learning has tended to ignore many of the other basic concepts in the rich philosophical tradition of semiotics. Indeed, the fundamental assumption behind the present work is that the study of semiotics provides a way of understanding what one might mean by saying that an infant has “learned a word.” This thesis will sketch a research program that will investigate whether human infants show evidence of learning the governing principles of semiotic systems.

In the current chapter of this thesis, some basic principles discussed in the study of semiotics are outlined. Then, these semiotic principles are recast as basic questions in the development of language. In the final chapter, an existing problem in language development is re-examined, and experimental evidence demonstrates how semiotic considerations may offer new theories of acquisition.

1.2 Semiotic Principles

Semiotics is relatively young, first recognized as a legitimate field of inquiry in the early 20th century, and primarily through the work of Saussure, Peirce, as well as through many others with whom they corresponded (Sebeok, 1994). In its modern incarnation, semiotics is in essence a structuralist field—that is, it assumes there are organizing principles which characterize how organisms (primarily humans) understand and use signs, and its goal is to describe these principles. In its modern context, semiotic approaches have been used in fields as diverse as geography, literary theory, and even biology. For the current purposes, however, only fundamental concepts from early writings are reviewed here, particularly those which might be applied to ask
psychological questions. First, social histories of the founders of this field are briefly outlined. This is then followed by a quick sketch of some of the basic, and largely overlapping, principles discussed by both Saussure and Peirce. Second, these principles are recast as psychologically relevant questions in language acquisition.

1.2.1 Saussure

Ferdinand de Saussure, lecturing on his ideas between 1906 and 1911, was a pioneer in linguistic thought. His contribution to the field of linguistics was paradigmatic: he redefined the very study of linguistics as a field that should focus on synchronic, or having to do with the static system of language, as opposed to diachronic study, or having to do with the process of evolutionary change. Indeed, the study of modern linguistics today-among both generative and functionalist approaches—is very different from what was taught not more than 50 years ago in most departments of the same name. This is due, in no small part, to Saussure. His thinking also contributed a great deal to the study of semiotics (or what Saussure called “semiology”). However, his description of language as a special case of a general system of signs has had less influence in linguistics. Perhaps because of this, the psychological claims that Saussure made in his writing are also less well-known, particularly in psychological approaches to language.

Much of the work published which is attributed to Saussure was actually printed posthumously, and furthermore was not written by himself, but rather compiled from the notes of two students who attended a series of lectures that Saussure delivered while at the University of Geneva. Their collection was published in 1916 with the title (translated from French), a Course in General Linguistics. This volume is considered Saussure’s
seminal contribution to the philosophical foundations of the field of linguistics. Due to the circumstances surrounding its publication, however, certain aspects of the arguments are, in some places, contradictory (see Holdcroft, 1991). For the purposes here, the gist of his arguments are outlined below, and then they will be recast in a more psychological light, particularly from the perspective of contemporary approaches in development.

1.2.2 Peirce

Charles S. Peirce developed his theory of signs between 1904 and 1913, mostly in correspondence with Lady Victoria Welby, an English aristocrat who was acquainted and interested in Peirce’s previous philosophical work, and was herself widely published in philosophical circles (Peirce, 1977). While never publishing a specific piece of work dedicated to his theory of signs, Peirce outlined his ideas in several letters to Welby, and in many other places among his collected writings on philosophical categories.

While Saussure drew most of his examples from natural language, Peirce was not primarily concerned with language as a descriptive case. Perhaps because of these different emphases, Peircean approaches are popular with those who have developed semiotics as a tool in the humanities (e.g., semiotics as it is used in literary theory, sociology, anthropology, etc.). However, many of Peirce’s early ideas map onto Saussure’s thoughts on natural language, despite the fact that, in all likelihood, neither had communicated with one another, nor were aware of published accounts of the other’s theory. Ultimately, both Saussure and Peirce recognized that language was only a subset of the study of semiotics, or a more general theory of signs.
1.2.3 Principles

The degree of similarity between the conclusions at which Saussure and Peirce independently arrived is astonishing. Both attacked previous philosophical interpretations of signs and words, and both made similar distinctions between the physical form of a sign, the referent of the sign, and the psychological link between the former two. In addition, there are several other similarities in the generalizations they made about how semiotic systems function. But because Saussure was primarily concerned with language, which is the focus of this thesis, basic Saussurian principles which govern sign systems are described below, and are annotated with philosophical contributions from Peirce.

Three main Saussurian principles governing the study of semiotics can be summarized in the following way: semiotics is the study of systems of arbitrary signs (Holdcroft, 1991). In this sentence, each bold term characterizes a principle governing semiotic systems. Using language as a descriptive example, Saussure defined these principles in the following way: language is composed of arbitrary relations between, to paraphrase Saussure, “sound-images” and “concepts” —“arbitrary” only in the sense that a signifier has no internal property related to the signified. These relational sets, or signs, are composed of bi-directional links between sounds and concepts, and at least in language, signs proceed in a linear fashion. Furthermore, these arbitrary signs occupy a place in a semiotic system, and thus are only defined negatively —that is, in relation to one another. Thus, arbitrary relations between sounds and concepts constitute a sign, and signs only have any communicative value when one recognizes that these signs operate in a linguistic system (Saussure, 1983).
Broad principles within Peirce’s theory of signs were not as explicitly stated as Saussure’s arguments in existing written accounts, but Peirce often alluded to central ideas from other areas of his work in philosophical categories. In this work, Peirce held that three main relations formed the basis of all other categories: Firstness, Secondness, and Thirdness. Firstness was generally meant to describe any first-order perceptions or sensations, like the sensation of perceiving a stone. Secondness described any relation between first-order perceptions—for example, the sensation of seeing that stone falling in relation to the ground. Thirdness describes a third-order relation, one which mentally relates perceptions of “secondness.” For example, being able to understand all situations involving falling rocks as instances of a gravitational principle might be understood as embodying Thirdness (Peirce, 1977). In linguistic cases, this relation is exemplified by a sign. For example, the sound /dog/ and the category “dog” depend on, or are mediated by, knowledge of English convention in understanding words. Thus, the relationship here is triadic in nature, where a third interpreting thought is required to link signs to that which they signify (Peirce, 1977).

These concepts—briefly sketched: the structure of signs, arbitrariness, systematicity, firstness, secondness, and thirdness—comprise some of the basic tools which are used by semioticians to study the use of signs in a wide variety of fields. The next section will target three of these basic concepts (i.e., arbitrariness, signs, and systems), expand on their philosophical implications, and discuss some psychological background related to these principles in order to generate research questions related to the field of word-learning.
1.3 Semiotic acquisition

In neither published accounts of Saussure's nor in those of Peirce's is much attention devoted to ontogenetic acquisition of the principles that organize semiotic systems. Their purpose, after all, was to characterize existing systems, not to build psychological theories of development. Hence, the purpose of this section is to act as a pointer. How can some psychological questions be motivated by a theory of signs? What sort of semiotically motivated questions can shed light on the process by which infants and children come to understand what "words" are actually doing? What follows are discussion of three semiotic principles recast in a more psychological light. Here, relevant evidence from existing work in developmental psychology is outlined in light of predictions made by semiotic theory.

1.3.1 Signs

As reviewed above, the critical distinctions made by semioticians in describing the structure of a sign—between a signifier, the signified, and the psychological / mental link between the former two—have more or less paralleled the way that word-learning researchers in infancy have tended to cluster their research questions. However, Saussure and Peirce generated more ideas about what constitutes a sign, and the purpose of this section is to discuss these ideas, as well as research in developmental psychology which suggests that there is development in understanding of signs. First, a discussion of the structure of signs is followed by a brief review of some existing literature from children and infants in the domain of pictures, maps, and scale-models, and then further by some speculative comments on whether that literature is comparable to research in language. Second, discussion of a second prediction derived from Saussure's and Peirce's
definitions of signs is followed by an experimental proposal, the results of which are detailed in Chapter 2.

Saussure defined a linguistic sign in the following way,

"...a two-sided psychological entity, [where a sound-image and a concept] are intimately tied, and each recalls the other." (Saussure, 1983)

Peirce, similarly defined a sign as,

"anything which is so determined by something else, called its Object, and so determines an effect upon a person, which effect I call its interpretant, and the latter is thereby mediately determined by the former" (Peirce, 1977)

In the first case, what might be notable about both of these passages is that Saussure and Peirce both posit a separation between the substance of a sign and its meaning. That is, the signifier itself is an independent entity from the concept or object which it denotes, but also crucially related in particular ways. What allows these two things (i.e., the signifier and the signified) to be related at all is psychologically determined, as Peirce called "the interpreting thought," and as Saussure later referred to as "form" determined by two "substances," but importantly a form which has psychological value. Thus, both assume three objects of study: the signifier, the signified, and the psychological understanding of the link between the two.

Indeed, a number of studies examining children's "symbolic understanding" have asked exactly this sort of question. When do infants and children understand that the signifier is independent from the signified, and separately "stands for" or "represents" something else? In the case of pictures and scale-models, infants and young children seem to experience difficulty in learning to distinguish between the signifier itself, and the real-world object which is depicted in the symbolic medium (DeLoache, 2005). For
example, although 5-month-old infants can discriminate between pictures and the objects which those pictures depict, infants continue to treat the pictures as real objects, scratching, sniffing, and otherwise trying to handle the 2-dimensional representations until 15-19 months of age (DeLoache, Strauss, & Maynard, 1979; DeLoache, Pierroutsakos, Uttal, Rosengren, & Gottlieb, 1998). Even by 20 months of age, children continue to make similar mistakes with miniature models of real objects; for example, until 24 months of age, toddlers continue to try to step into miniature models of cars when initially given larger, child-sized versions to play with (DeLoache, Uttal, & Rosengren, 2004). These results suggest that learning a distinction between signifiers and signifieds is difficult for infants and children, particularly when the two are so perceptually similar, as in the cases examined by DeLoache and colleagues. At a minimum, this work suggests that development of the concept sign is certainly non-trivial for infants and young children.

But how does this work on “symbolic understanding” translate to the domain of language? Certainly, children understand that words are separate from the concepts which they denote by 20 months of age, but is there developmental progression in this domain as in the domain of scale-models and pictures? What kinds of evidence would be needed to show progress in comprehending linguistic signs? To predict patterns of development, an appeal is made to Peirce’s categorical distinctions between firstness, secondness, and thirdness. As Peirce noted, the philosophical distinctions mentioned above were meant to be universal ones and relevant at every level of any conceptual hierarchy. Work from DeLoahe and colleagues suggest, as well, that development can also be seen in infants’ understanding that there is indeed a functional difference between signifiers and
signifieds. This exactly the spirit of the question asked here: can children have different degrees of understanding in grasping the distinction between a signifier and a signified in the language domain? Might there be a way of characterizing these degrees of understanding, particularly one which is guided by philosophical distinctions, such as those corresponding to Peirce’s philosophical categories? Some evidence suggests that, in the case of linking speech and concepts, this is indeed the case.

As Gogate and colleagues have argued, infants’ early perceptual abilities may be initially limited in the following sense: infants seem initially to look only for amodal correspondences between sounds and objects in their environment. Only at older ages do infants seem to learn to link speech sounds in an intermodal (i.e., synchronous, but arbitrary) fashion, learning to remember co-occurrences between speech and objects by 8 months of age, and only later are able to learn completely arbitrary relations (Gogate & Bahrick, 1998; Gogate & Bahrick, 2001). Cross-modal matching studies have shown that infants as young as 2 to 4 months perceive many amodal properties: learning quickly how visual texture is related to tactile texture (Meltzoff, 1979) and how sounds are related to movement and collision of different objects (Starkey, Spelke, & Gelman, 1990). Infants may also have an equivalent understanding of speech: 2- and 4-month-old infants look preferentially to a matching face sound when a corresponding speech stream is played over speakers (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999; Patterson & Werker, 2003). One might conclude, in the case of speech as well as other domains, that by 2-4 months of age infants demonstrate “firstness” between sensing one stimulus and sensing another (often in another modality), yet not drawing a clear functional distinction between the two. In this sense neither signifiers, signifieds, nor significations are
distinguished. Evidence which suggests that infants can learn amodal, but not intermodal correspondences in early infancy supports this preliminary conclusion.

However, are intermodal correspondences the only way in which infants may link speech and objects in early infancy? Infants at 6 to 12 months of age may instead begin to perceive speech sounds as being related to concepts in a way which is neither entirely intermodal nor referential in nature. For example, when 6 to 12-month-old infants hear a word presented in the presence of a static picture, this may be enough to categorize the pictures as a basic-level concepts without generating a referential link (Balaban & Waxman, 1997; Waxman & Markow, 1995), and hearing two different words may act as an index for the number of sortal kinds present, but not currently in view (Dewar & Xu, under review; Xu, Cote, & Baker, 2005). These kinds of findings suggest that speech and concepts are initially linked, but not necessarily in just a low-level Gestalt-related perceptual fashion. This kind of link may be equivalent to Peirce’s notion of secondness; infants at this age begin to perceive words as being separate entities from the concepts they label, but not yet as full-blown referential signs. Rather, words may act to direct attention to concepts in a non-referential way, via secondness in the structure of signs. In this sense, infants may understand the difference between signifiers and signifieds, but not the psychological aspects of referentially linking the two.

One possibility is that this kind of proto-sign might have the form of what Peirce called a “sinsign,” where the signifier’s relation with the signified is dependent on its contextual relation to it (Peirce, 1977). Similar to the way that, to an adult, an extended finger pointing into a box indexes a particular object, this type of sign is not strictly referential in nature, and is dependent on its relation to the referred object in a context-
specific way (e.g., which direction the finger is pointing). Distinct words in the same context may indicate the presence of distinct conceptual categories in that context, similar to the way that distinct points unambiguously refer to distinct objects. This hypothesis differs subtly from that offered by Xu (2002) and Waxman (2004) in the sense that words, at this age at least, are meaningless independent of a real-world context. For example, just as distinct words leads infants to predict the existence of distinct kinds of objects (Dewar & Xu, under review), infants hearing the *same* two distinct words in another context would again expect to see different kinds, but hold no expectation to see the same two kinds as in the first context.

Infants learn words' symbolic value in by at least 18 months of age (Akhtar & Tomasello, 1996; Baldwin, Greene, Plank, & Branch, 1996; Bloom, 2004; Hollich et al., 2000). However, as early as 12-15 months of age, infants show some learning of new words in more referential tasks: for example, preferentially looking at particular objects in choice tasks and extending a label to similar objects (Golinkoff et al., 1994; Hollich et al., 2000). These studies, at least, suggest that infants have the capacity to learn *reference* by 12-13 months, at least under optimal conditions. If further work supports this conclusion, it would indicate that infants have thus acquired a fully triadic relationship: words, interpreting thoughts, and referents. However, this conclusion is perhaps controversial, as some studies have suggested that infants at this age more readily understand that words have a "goes with" rather than a "stands for" relationship with their referents (Werker et al., 1998; Werker & Tees, 1999). However, several pieces of evidence suggest that, even though word-learning is far from easy by the end of the first-year of life, infants at least *can* show some degree of maturing and non-adult-like.
referential competence. Hirsch-Pasek and colleagues suggest that 14-month-old infants show some evidence of extending novel labels to objects of different colors (Hirsh-Pasek, Golinkoff, Hennon, & Maguire, 2004). Woodward suggested that 12- and 13-month-old infants are sensitive to eye-gaze as an intentional cue (Woodward, 2003), and furthermore selectively used these cues in word-learning tasks (Woodward, Markman, & Fitzsimmons, 1994; Woodward, 2004). Moreover, electrophysiological indices of reference do show development from 13- to 20-months, but an overall review of the literature suggests continuous, rather than discontinuous change (Mills, Coffey-Corina, & Neville, 1997; Mills et al., 2004; Mills, Conboy, & Paton, 2005). However, this does not say that further development is not necessary, or that infants do not also learn a “goes with” relationship between words and concepts in some tasks (Schafer & Plunkett, 1998; Werker et al., 1998). Indeed, the literature on word-learning is complex and nuanced on this point. However, the evidence above still suggests that some word-learning tasks succeed in teaching referential signs by at least 12-months of age. As will be suggested in the section on “arbitrariness” below, Peirce’s categories not only characterize development in the degree to which there is a distinction between the components of signs, but also the type of relationship which underlies them.

In an interim summary, it has been argued that a) signs have a tripartite structure, b) that in the domain of pictures and scale-models, infants and children show development in learning the distinction between signifiers and signifieds, and c) that in the domain of language, Peirceian distinctions may further characterize development of this tripartite structure of words.
A second prediction follows from Saussure and Peirce’s notion of how signs are defined. In Saussure’s construal, looking back at the definitions outlined above, the “concept” (i.e., the idea denoted by a word) and the “sound-image” (i.e., the word form itself) may recall each other. Saussure suggests that signs are “two-sided” entities, where sounds invoke concepts, and therefore concepts also invoke sounds. In other words, signs are bi-directional. Peirce, on the other hand, suggests that a sign acts as a “mediator” between one’s own “interpreting thought,” and an “Object.” While not explicitly equivalent to Saussure’s notion of bi-directionality, the determining relationships stated in Peirce’s definition are implicitly relevant to this idea. To illustrate this, Peirce’s notion of understanding a sign is paraphrased again here: some object “determines” the use of a particular signifier, which in turn “determines” a psychological state (i.e., the interpreting thought). To invoke an example used by Hookway (1985), a deer might break off a piece of bark from a tree, and thus produce a sign. Upon seeing the bark, a hunter might make an interpreting link between broken bark and deer. Essentially, the presence of the deer is related to the hunter via the signifier: the broken bark. Hookway (1985) suggests that the causal direction can follow in the other direction, as might be common in speech production. If the hunter wished to inform her fellow hunting-group of a deer’s presence, she might utter the word, “deer!” In the reverse direction, the hunter’s interpreting thoughts about the deer would determine her uttered sign, which would in turn act upon others’ interpreting thoughts within the hunting-group, and bring to mind the concept of a deer.

The question here is: how do often abstruse, philosophical notions translate into questions of psychological interest? In the case of Saussure, the distilled expectation
seems simple: sounds should influence mental representations of concepts, and concepts should also influence mental representations of sounds. In Peirce’s case, it seems a bit more complicated. One important contribution from Peirce concerns the objects-of-study in thinking about how signs operate. According to Peirce (and perhaps not as Saussure might originally have formulated the argument), the substance of the sign (i.e., the signifier) is always a mediating entity, and strictly acts either on an “interpreting” psychological representation, or on a conceptual “Object” of the sign. The take-home message might consist of something like the following: unlike what Saussure suggests, it is not “sound-images” and “concepts” which are in a bi-directional relationship, but rather the “concepts” and “interpreting psychological representations of signs” –or, in other words, it’s not about the sounds themselves, but rather the psychological construal of those sounds as signs.

This discussion of the finer points of Peirce’s philosophical arguments may not be critically relevant, in the end, to the question of bi-directionality. However, it might be important to think about how whether or not distinctions between phonetic (i.e., acoustic) versus phonological (i.e., meaning-oriented) processing has its roots in these Peirceian distinctions. Results from one psychological study which was designed to test the bi-directionality hypothesis are outlined in the last section of this thesis. First, however, two more semiotic principles are discussed in a psychological context.

1.3.2 Arbitrariness

Saussure suggested that the sounds of words are arbitrarily related to the concepts which they denote. In other words, nothing about the phonetic quality of sounds used in a word is related to the concept that the word denotes, and furthermore, any sound could
arbitrarily be substituted for a concept if a community of speakers agreed on such a change. In the domain of speech, Saussure noted two exceptions to this rule: first, cases of onomatopoeia, where the phonetic qualities of a word (e.g., *meow*) are not arbitrarily related to the referent (e.g., sounds which cats actually make), and second, in the cases of interjections, where an utterance (e.g., *ouch* or *wow*) might have at least some sort of non-arbitrary, or conspecific quality. Saussure devoted little attention to these alternatives, and instead briefly suggested two counter-arguments: first, that these cases are on the periphery of what is considered modal use of language, and second, that both onomatopoeic words and interjections are, to a large degree, language-specific (and in some sense arbitrary), and subject to the same kinds of evolutionary changes that govern sound-change (Saussure, 1983). In a later portion of *Course in General Linguistics*, Saussure recognizes that some words are “motivated” by inflectional processes, where rule-governed structure dictates how certain morphemes can combine to yield words.

Momentarily allowing Saussure the benefit of the doubt on these exceptions, one psychological question of interest is then: *must humans learn this principle in the course of learning words?* Indeed, little experimental work has directly suggested that even adults have such a principle. Recently, however, some studies have been suggestive of this fact. Westbury (2005) asked adults to perform lexical decisions for words and non-words which were presented visually on a screen and either embedded in a black-on-white “spiky” or “curvy” frame. He also manipulated the phonetic quality of the words and non-words themselves, using words with only obstruent stop consonants (i.e., “toad” and “kide”) or only consonants that were also continuants (e.g., “lole” and “moon”). Drawing from Gestalt theory, Köhler (1947) originally suggested that perception of
discontinuous speech sounds (i.e., obstruents) may be related to perception of discontinuous visual contours (e.g., “spikes”), while conversely, continuous speech sounds (i.e., continuants) may be related to perception of continuous visual counters (e.g., “curves”). However, Westbury found that adults’ latency to correctly identify known words did not show an interaction between visual frame and phonetic quality. However, adults’ latency to reject non-words showed a Gestalt-like pattern, where adults were faster to reject continuant-containing words in curvy frames and obstruent-containing words in spiky frames than the converse examples. Thus, although adults were sensitive to these perceptual correspondences, and the visual properties of the frame helped adults process and then successfully reject non-words, a Saussurian principle correctly predicts that adults would not be influenced by Gestalt correspondences when processing known words (since only words have linguistic status as signs). However, given that this study was evaluating processing, it remains unanswered whether or not the learning of words is affected by these principles in adulthood.

Interestingly, studies with both adults and young children have suggested that humans do use this sort of Gestalt information to match referents with novel word forms. Maurer, Pathman, & Mondloch (2006) suggested that both adults and children 2.5-3 years of age, given two possible shapes, selected the Gestalt-related shape (e.g., a “spiky” or a “round” one) when hearing words with either more obstruents or more continuants (e.g., the shape called “keiki” or “mouma”). Interestingly, however, Smith & Sera (1992) suggested that Gestalt-like pattern matching is affected adversely by learning linguistic terms, at least in some cases. In their studies, 2-year-old children succeeded in matching darker gray mice with larger models, and lighter gray mice with smaller ones. However,
both 3-year-olds and adults did not continue to show similar matching patterns, presumably because the older children and adults had learned linguistic terms “dark” and “light.” This pattern, however, did not hold when these groups were tested with other kinds of Gestalt correspondences (i.e., big-small matched with loud-soft).

These developmental studies still suffer, however, from the same problem that plagues other adult work which purports to refute claims of arbitrariness (Brown, Black, & Horowitz, 1955; Taylor & Taylor, 1962; Taylor, 1963). In typical experimental studies of this kind, adults are asked to determine whether words in evolutionarily distant languages (e.g., English speakers hearing Chinese or Czech words (Brown et al., 1955)) or artificially created ones (Johnson, Suzuki, & Olds, 1964) correspond to either one of two categories contrasting on some semantic continuum: light or dark, heavy or light, good or bad, etc. Usually adult performance is above chance in these studies, matching the pair of foreign words to the correct semantic pole (see Nuckolls, 1999 for review). Presumably, adults are depending on Gestalt-like strategies to guide their decisions: for example, sonorants (i.e., vowels and certain consonants) with lower formant frequencies are linked to darkness, heaviness, roundness, etc., while sonorants with higher formant frequencies are linked to brightness, lightness, sharpness, etc.

At first blush, as in the developmental studies previously mentioned, this might be taken as evidence against such a “principle of arbitrariness” — that is, sometimes adults and children seem to show a preference to link a word form to an object with appropriate Gestalt-related properties. However, in every case — including both in adult and developmental studies — subjects are asked to choose a referent from two possible options without the benefit of any unambiguous referential information. This kind of task is not a
strong test of a principle of arbitrariness, because in the absence of any explicit cues for reference, subjects’ decision processes may lead one to rely on other, non-semiotic heuristics in eventually making a Gestalt-related choice. From the view of the naïve adult, Köhler and Saussure can both be right — arbitrariness is the rule when given explicit evidence that a label unambiguously refers to a particular referent, but in the absence of such information, one may rely on Gestalt strategies.

A truer test of the notion that adults expect newly generated signs to be arbitrary is to run a teaching study. In such a study, both adults and children might be taught a word, and then evaluated on ease of access to that word in memory. Critically, competence in learning the word would be assessed between items, such that adults and children would be taught to link words and shapes, some pairs of which would be related in Gestalt-specified ways and other pairs in the opposite manner. In contrast to the previous studies, subjects would not be asked which object a speech sound might denote without explicit referential cues, and then tested on their decision processes. Rather, subjects would be assessed on memory for and access to the link across each word-object pair. If Saussure is correct, at least in the psychological aspects of the arbitrariness claim, then memory for the word-object link should be equivalent regardless of the relation between its phonetic and physical characteristics.

The fact that studies showing adults can often “guess” the semantic polarity of words from unrelated natural languages, however, remains a problem for Saussure’s claim. Phrased another way, even if adults were relying on non-semiotic processes to arrive at their decision, does it not speak against Saussure’s claim that adult intuitions correspond to the actual words used in the languages tested? Indeed, the process of sound
symbolism is a topic that is well studied (Hinton, Nichols, & Ohala, 1995). These linguists have suggested that several linguistic phenomena rely on non-arbitrary relations between semantic properties and phonetic segments: correspondences between vowel height and semantic size, articulatory manner corresponding to verbal properties, etc. (Hamano, 1998; Hinton et al., 1995). While Saussure would probably be skeptical that such phenomena would be part of the regular linguistic system and resist sound-change over the evolution of any given language (thus rendering the sounds essentially arbitrary), the present state of the psychological literature is agnostic on this point. The fact remains that too little experimental work has been carried out on these phenomena.

Of interest, however, is whether there may be any developmental trends in adults’ versus children’s performance in these modified teaching tasks. If there is development in the principle of arbitrariness, what kinds of observations might be characteristic of a developmental trend? Again, semiotic theory may be of some help here. Peirce’s categorical distinctions between firstness, secondness, and thirdness were meant to identify fundamental distinctions which might be applied universally — here in the case of relating signifiers and signifieds. Notably, the use of Peirce’s terms in this current discussion on the arbitrariness is subtly different from the use of these terms in the previous discussion on signs. In the previous case the terms for firstness, secondness, and thirdness characterized different levels of understanding that signifiers and signifieds were in fact, independent of each other. In this current view, it is stipulated that signifiers and signifieds stand in relation to one another via a linking relationship. At stake is what kind of referential relationship this is. Peirce drew a distinction between three types of signs: icons, indices, and symbols, which embodied his philosophical categories of
firstness, secondness, and thirdness, respectively. In the first case, signs such as perfumes, pictures, and maps are examples of icons – that is, the internal structure of the signifier resembles or simulates the real-world concept which it denotes. In the second case, indices include signs such as weathervanes, finger-pointing, a broken twig (when hunting deer), where the intrinsic properties of the signifier (e.g., its proximate cause, or its relative position in time and space) is related to the signified. Peirce’s third case, of course, is equivalent to Saussure’s characterization of arbitrariness, where symbols are any kind of socially, arbitrarily, and conventionally determined sign, of which words are a prime example.

One possibility, of course, is that infants and children will show developmental parallels in perceiving speech sounds in-step with Peirce’s categories. For example, as infants begin to learn words and are old enough to know that these words are signs which stand in a referential relationship to concepts, they may still perform better in tasks where words are linked with referents in a Gestalt-specific way –at least initially. That is to say, children may preferentially learn words which have onomatopoeic qualities to them, and prefer words which act as icons, simulating the properties of their referents. For example, there seems to be anecdotal evidence that children would preferentially learn the phrase “woof-woof” rather than the word “dog” specifically to refer to a child’s pet. Whether or not young children are better able to learn words of this sort under experimental conditions, of course, is an empirical question. Given that infants first succeed at mapping arbitrarily related novel words to novel referents at 12-15 months of age in what are argued to be referential tasks (Hollich et al., 2000), one prediction is that even younger infants will be able to map an onomatopoeic word in the same manner.
How would words which have the properties of a Peirceian “index” be characterized? As mentioned above, Saussure made a distinction between words which are “motivated” in comparison to other words in the lexicon: this may include words have more regular inflectional properties, and thus have predictive phonetic but not, importantly, onomatopoeic qualities. For example, a child might have an easier time learning that the suffix “-saur” might refer to any member of the super-ordinate category “dinosaur” because the phonetic sequence co-occurs with, but does not simulate, other similar kinds of referents. While the above example is concerned with children’s learning of morphologically regular nouns, this point is reminiscent of the claim that young children tend to over-regularize past-tense forms of verbs in development (Marcus, Pinker, Ullman, & Hollander, 1992). Furthermore, some languages, like Japanese, have whole systems of “mimetic” words, which are characterized by correspondences between some semantic and phonetic characteristics (Hamano, 1998)). Interestingly, Nagumo, Imai, and colleagues (2006) suggested that younger children were able to correctly generalize verbs when these verbs seemed to have a mimetic relation with a novel action, compared to when this relationship was not mimetic.

In summary, there is some evidence that when children and adults are faced with an ambiguous choice about what a word labels, their decisions are not entirely arbitrary in nature. However, it is unknown whether adults’ learning of words in an unambiguous context would allow for equivalent retention and memory for arbitrary and non-arbitrary word-object pairs. Some evidence suggests that children tend to learn words which are less arbitrary better than words which have a fully arbitrary relationship with their referents (Nagumo et al., 2006). Critically, however, a Saussurian prediction suggests that
any advantage in the child’s learning of Gestalt or onomatopoeic correspondences between words and objects should disappear in adulthood.

1.3.3 Systems

Saussure was fond of drawing analogies between language and the game of chess. Some of these analogies were related to Saussure’s notions about signs: for example, a knight in form and material is meaningless outside of the game itself, similar to how a signifier is only imbued with linguistic value if it is treated as a sign. Similarly, the use of a knight with its specific shape and name are conventional and arbitrary; any piece could be substituted for that knight if any community of players agreed on the substitution. One further analogy with chess is the notion that pieces are only considered significant in relation to other pieces’ current position on the board—an allusion to his argument that language should be studied as a synchronic system. In a sense, this claim is also an attack on the nominalist tradition in philosophy, which suggested that the kind of sounds and concepts in the world were pre-packaged, and the challenge of language was to understand how these two things are linked. Rather, Sassure (1983) was a relativist to the extreme:

“Without language, thought is a vague, uncharted nebula. There are no pre-existing ideas, and nothing is distinct before the appearance of language. Against the floating realm of thought, would sounds by themselves yield predelimited entities? No more so than ideas .... [phonic substance is a] plastic substance divided in turn into distinct parts to furnish the signifiers needed by thought.” (p. 112)

This position is certainly one which was at the center of Saussure’s argument. Language is a system, where the “value” of a sign is determined only in relation to other
signs within the same system. However, current views in both speech perception and
cognition are certainly not as radical as the strong version of Saussure's view,
exemplified in the excerpt above. For example, there is plenty of articulatory and
psychophysical evidence for natural and universal discontinuities in the perception of
sounds (Kluender, Diehl, & Killeen, 1987; Lisker & Abramson, 1964), and further
evidence that these distinctions occur without any corresponding meaningful linguistic
contrasts (Best, 1988; Streeter, 1976). In the domain of concepts, Quine (1960) not only
agreed with Saussure's position, but argued eloquently for it; however, evidence
nevertheless suggests the existence of some pre-verbal ontological divisions (Soja, Carey,
& Spelke, 1991).

How, then, might Saussure's prediction cash-out in a psychologically relevant
way? One possibility is to test a weaker version of this hypothesis: that some phonetic
and conceptual distinctions are universal, but in cases where things are more ambiguous,
categorization strategies may differ across domains. Two questions are related to this
issue. First, is Saussure fundamentally correct in drawing a distinction between
categorization related to semiotic systems, as opposed to non-semiotic systems? The
following type of evidence would support this claim: if categorization differed when
elements in some medium are perceived as meaningful signifiers or signifieds (e.g.,
phonemes in speech, or animals categories in cognition, etc.), compared to cases when
those elements are perceived outside of a semiotic system (e.g., vocal sounds, shapeless
blobs). If this were indeed the case, a second question related to Saussure's hypothesis
becomes relevant: is the medium of both signifiers and signifieds equally affected by
such semiotically driven differences? Evidence supporting this second assertion might
take the following form: if categorization strategies in the medium of signifiers (e.g., speech) differed from categorization strategies in the medium of signifieds (e.g., artifacts, natural kinds, visual shapes, etc.)

Indeed, this proposal is reminiscent of some previous work in speech perception. Drawing from the work of Trubetskoy and other linguists of the Prague School, such work suggests that, unless some phonetic contrasts represent a meaningful distinction for listeners in their native-language, they remain difficult to discriminate (Best, 1994; Polka & Werker, 1994; Werker & Tees, 1984a; Werker, 1995). However, evidence of this sort is only equivocally related to the above hypothesis regarding different categorization strategies in semiotic-versus non-semiotic systems. This is because some meaningful phonetic distinctions are often correlated with the statistical or distributional characteristics of the input (Maye, Weiss, & Aslin, in press; Werker et al., in press). In these cases, a single categorization strategy may allow subjects to discriminate a native-language phonetic contrast, simply because they receive statistical evidence for this distinction, and would perform well on discriminating a non-native contrasts given adequate statistical input. Indeed, this is similar to categorization proposals in the conceptual domain, which rely on statistical input to shape infants’ early notions of basic-level categories (Eimas & Quinn, 1994). Future work will need to distinguish these possibilities.

In summary, the third semiotic principle discussed here concerns systems of signs. As Saussure suggested, value in a sign system is defined negatively, and only in relation to other signs. Thus, categorization of the substance of either signifiers or signifieds may differ when perceived as being a part of sign systems, as opposed to when perceived as
existing outside of a sign system. If this were indeed the case, further work would be needed to show whether this were true for the substance of both signifiers and signifieds.

1.4 Concluding Summary

This section has attempted to discuss some general ways in which philosophical thinking about a theory of signs can inform work in developmental psychology. In particular, at stake is whether infants must learn the principles which guide communicative systems: primarily, the system of language. Three general principles were elaborated, and some ideas of what might constitute evidence for these principles were suggested. Table 1 summarizes some of these suggestions.

In the following section, an empirical study focuses on a particular aspect of the theory outlined in Table 1. In particular, the work below examines whether or not infants show evidence of having an intermediate understanding what the composition of signs are. At issue is whether infants 6 to 12 months of age are able to link signifiers (i.e., word forms) with signifieds (i.e., objects), but not necessarily in a relationship equivalent to that of a "word" (i.e., lacking the appropriate interpretation relation). Some of the evidence for this intermediate link is reviewed, and one other semiotic prediction is directly tested. Is this relationship between a speech sound (i.e., a signifier) and a concept (i.e., an object) bi-directional? In other words, two interrelated questions are posed: to what degree do word forms influence the structure of concepts, and to what degree to concepts influence the structure of word forms?
Table 1: Conceptual summary for psychological applications of semiotic theory.

<table>
<thead>
<tr>
<th>Principles</th>
<th>Theoretical</th>
<th>Empirical Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Signs</strong></td>
<td>- Signifier related to signified concept via interpreting thought</td>
<td>Infants:</td>
</tr>
<tr>
<td></td>
<td>- Signifier and signified concept are bi-directionally related</td>
<td>a) Fail to understand difference between signifier and signified concept (2 - 4m)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) Know that signifiers are separate, but related entities in relation to signifieds; fail to understand interpreting thought (6 -12m)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c) Signifiers come to “stand for” signified concepts in an interdependent fashion within a communicative system (14m – onward)</td>
</tr>
<tr>
<td><strong>Arbitrariness</strong></td>
<td>- Distinction between icon, index, and symbol</td>
<td>Infants and children incrementally learn icons, then indices, then symbols more easily</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adults learn all equally well, when clearly operating within a communicative system</td>
</tr>
<tr>
<td><strong>Systems</strong></td>
<td>- Categories only defined in relation to each other within semiotic systems</td>
<td>Categorization might differ in semiotic versus non-semiotic mediums</td>
</tr>
</tbody>
</table>
2. Empirical work on signs

2.1 Introduction

Since Saussure, semioticians have characterized spoken words as prototypical examples of a semiotic sign, consisting of sounds and concepts linked in a signifying relation. Presently, it is asked whether semiotic interpretations of how words function as signs might be interpreted from a developmental perspective. What is the nature of an early link between a speech sound and a concept, and how might this relation be characterized? Do young infants show an emerging notion of what a sign, and hence a word, might be?

Presently, it is suggested that infants show development in understanding the semiotic character of words. Evidence is reviewed below which suggests that infants 6 to 12 months of age seem to know that sounds and concepts are related, but not necessarily in a referential way. Rather, some evidence suggests that infants are not able to robustly link sounds and concepts in a referential manner until the end of the first-year of life. The nature of this early link between speech sounds and concepts is further explored in the present set of studies.

Specifically, it is suggested that this early link between sounds and concepts is bidirectional, and allows young infants to do two things: first, as previous research already suggests, hearing a word in a labeling context may influence infants' perceptions of how their conceptual world is organized. Second, the converse may also be true: hearing words in a labeling context may tune infants' perceptual systems, allowing them to successfully identify, encode, and remember word forms. This second possibility is tested here.
In summary, two types of literatures on infants’ learning of early words are particularly relevant for the argument made here, and are briefly reviewed below. One set of studies characterizes the link between sounds and concepts in one direction, suggesting that words can influence how infants perceive and remember objects. This includes studies with both younger (6-12 months of age) and older (12-18 months of age) infants. Studies with younger infants have suggested that words act to categorize, individuate, and predict the existence of conceptual categories. At older ages, infants are further able to map words onto their referents, but this ability shows development in both the 1\textsuperscript{st} and 2\textsuperscript{nd} year of life. And while these studies suggest that speech sounds in a labeling context influence and point to infants’ conceptual categories, another set of conversely related studies suggest, instead, that the direction of the link can also proceed in the other direction. Studies of this type with younger infants would investigate categorization and identification of word forms, and is exactly the topic of the present empirical work. Related studies with older infants are also reviewed below, which suggests that infants show development in their ability to remember and retrieve word forms. A conceptual summary of type of studies mentioned below is illustrated in Table 2.
Table 2: Conceptual summary of studies in early word-learning

<table>
<thead>
<tr>
<th>Directionality</th>
<th>Under ~12 months</th>
<th>Over ~12/14 months</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sounds -&gt; Concepts</strong></td>
<td>Words act to categorize, individuate, and predict the existence of different conceptual categories.</td>
<td>Distinct words begin to “refer to” specific kinds of things.</td>
</tr>
<tr>
<td><strong>Concepts -&gt; Sounds</strong></td>
<td>Objects (i.e., concepts) may act to categorize word forms as distinct from one another.</td>
<td>Distinct objects are referred to with word forms contrasting on native-language phonemes.</td>
</tr>
</tbody>
</table>

2.1.1 Word-learning at 12 months and beyond

At the onset of “word-learning proper,” studies suggest that infants as young as 12 months of age are just beginning to learn that words refer to particular kinds of objects, but only with heavy use of pragmatic, social, and intentional cues. However, learning is still affected by irrelevant attentional dimensions, like the perceptual saliency (i.e., “attractiveness”) of an object itself (Golinkoff et al., 1994; Hollich et al., 2000). Yet, preliminary evidence suggests that infants as young as 13 months of age selectively learn words when they are presented in a socially relevant context—that is, they will learn words when a speaker is looking at an object, but not when she is looking at a video screen and producing the word (Baldwin et al., 1996; Woodward, 2004). These data suggest that infants may consider the intentions of a speaker when deciding how to link objects with words. However, some evidence suggests that words have a less referential relationship with concepts at this age. Werker et al. (1998) suggest that infants can learn to discriminate changes in links between novel words and novel objects independently of
any perceptual (e.g., synchrony) or social-intentional cues at 14 months of age. From there, infants must continue to learn how to use social, pragmatic, phonological, and attentional cues before being able to accurately and quickly learn the meanings of words (Baldwin & Moses, 2001; Werker & Curtin, 2005). Thus to say that at “older” ages, infants begin to map words referentially is by no means suggesting that word-learning is “ready” at 12 months. Rather, learning shows development such that, by 18-months of age, infants become highly proficient word-learners (Hollich, Newman, & Jusczyk, 2005; Werker & Tees, 1999; Werker & Curtin, 2005). This is particularly uncontroversial.

Notably, however, two studies seem to contest the assertion that infants do not learn specific referents before 12 months of age. As early as 6 months of age, infants look preferentially at photographs of a parent’s face over a gender-matched stranger’s face when hearing the word Mommy or Daddy (Tincoff & Jusczyk, 1999). These results are especially striking, because the method is similar to other studies which have used preferential looking to evaluate infants’ knowledge of links between referents and words (Hollich et al., 2005). One general problem with this type of study, however, may be related to the fact that the words “Mommy” and “Daddy” and their referents are ubiquitous in English-learning infants’ environment. Indeed, it remains to be seen whether this particular example is anything like word-learning at older ages: for example, do infants understand that the word refers to the entity “Mommy,” and not to any correlated but independent properties in the world (e.g., a particular scent, milk, warmth, general comfort, etc.). Furthermore, it seems strange that these early “words” are proper nouns, whereas other research suggests that infants have an initial bias to interpret words
as count nouns, learning proper nouns in a laboratory setting only at older ages (Hall, Lee, & Belanger, 2001).

In another set of studies which tests children's ability to learn links between novel word forms and novel objects, Gogate and colleagues suggest that by 8 months of age, infants are able to learn to link words and objects (Gogate & Bahrick, 1998; Gogate & Bahrick, 2001). Infants dishabituated to pairings between novel word forms and videos of novel objects, but only when the word and movement of the object were synchronous. But, again, because learning the link was only successful when stimuli were synchronized, detection of this link might be due to infants' perception of the words as an intermodal property of the object itself, rather than a separate, referential label.

In summary, evidence that children are able to learn the specific referents of words prior to the first-year of life is sparse, and seems subject to qualifications about the type and nature of word-learning. Rather, infants seem better able to recognize and manipulate the specific referents of word labels by 12 months of age.

Several studies also studying infants over 12 months of age ask, instead, how infants show development in learning and recognizing word forms (Naigles, 2002; Werker & Curtin, 2005). Studies of this type usually investigate situations where distinguishing the referents of words is relatively easy, but selecting among minimally different and familiar word forms (e.g., "baby" and "vaby") is relatively difficult. These similar labels are often treated as distinct by older infants, in both recognition and learning tasks. For example, in word recognition studies of this kind, infants 14 months of age look preferentially at a familiar object when hearing its label, but not when hearing a slightly mispronounced version of that word (Swingley & Aslin, 2000; Swingley &
Aslin, 2002). In word *learning* studies of this type, however, the evidence is more complicated. In a typical study investigating learning of minimally different word forms, infants are habituated to either one, or two pairings between objects and words. In a test phase following a single-pairing habituation, infants hear a new word paired with the same object; in a test phase following a two-pairing habituation, infants see the same objects and hear the same words, but the *pairings* are switched. Thus, dishabituation in these types of studies suggests that infants have encoded the pairing(s) between the word(s) and the object(s) (Werker et al., 1998). Learning minimally contrastive word forms in this manner is difficult for infants: 14-month-old infants are able notice the switch in word-object pairings, but not when minimally different word forms are used, despite the fact that they may still be able to discriminate the word forms in a non-object version of this task (Stager & Werker, 1997). However, as more familiar word forms or objects are used, or as experimental demands are reduced, infants are able to learn, or at least able to *encode* these minimally different novel words (Fennell & Werker, 2003; Yoshida, Fennell, Swingley, & Werker, 2006). To summarize, by at least 14 months of age, distinct objects can be *encoded* with to similar-sounding word forms, but remembering and retrieving these links is still difficult until older ages (Fennell & Werker, 2003). Improvement in memory and retrieval for phonetically detailed word forms may be due to the development of qualitatively new representations for word forms (Werker & Curtin, 2005).

In summary, both types of word-learning studies suggest that beginning around 12 to 14 months of age, infants learn that distinct words can be linked (and may also refer to) specific objects or *kinds* of objects, and conversely that distinct objects or kinds can be
linked to similar-sounding, but nevertheless distinct word forms. However, infants must still learn how to efficiently make this mapping: in the former case, learning how to use various social-pragmatic, intentional, and attentional cues to learn word referents, and in the latter case, developing processing strategies to remember phonetic detail in newly learned word forms.

Of particular interest, then, is the question of what are early words; what do young infants, between 6 to 12 months of age, expect words to do? One possibility, of course, is that words are functioning to form categories and concepts which might be useful in later comprehension of language, particularly learning specific word forms and specific kinds of things. For example, older infants' first assumption about what words might label is, of course, kinds of things—only later do they learn that some types of words may also denote particular individuals (Hall et al., 2001). How then, do these infants learn what kinds are?

In a similar fashion, the complementary question of how word forms are learned may also be asked. For example, what defines the notion of "minimally different" or even "possible" word forms? Indeed, some preliminary evidence suggests that, even for infants as old as 17 months of age, word forms must be composed of segments drawn from one's native-language inventory (Dietrich, 2006). But this only raises the related question, where do these categories come from? Following in the tradition of Saussure, Quine, and other philosophers, this may suggest that, at least to some extent, the categories used in language may be formed as infants begin to learn words. The question, then, is how?

One possibility is that infants first learn that words have some unspecified kind of communicative function, which helps set-up categories useful for learning specific types
of referents and specific word forms. Stated in another way, there may be an intermediate step before infants can learn to link words and specific concepts. Infants may know that some speech sounds and concepts are related, yet do not fully understand the exact communicative function of words. Instead, words may act as “pointers” for signaling the existence of unique concepts, and concepts may similar act as “pointers” to signal the existence of distinct words. There are, then, two possible types of studies which are relevant for this point: first, studies which focus on “words-as-cues-for-concepts,” and second, the current study which focus on “concepts-as-cues-for-words.”

2.1.2 Words-as-cues-for-concepts

Several studies suggest that, beginning around 6 months and continuing into the 1st year of life, human infants develop an evolving understanding of the communicative significance of words. For example, Fulkerson & Waxman (2006) report that 6-month-old infants are able to form basic-level category representations (e.g., dinosaur) when familiarized to a series of pictures of category exemplars (e.g., different kinds of dinosaurs). Evidence for this comes from infants’ looking patterns in a test phase displayed after the familiarization phase: when given the choice to look either at a picture of a novel exemplar within the category (e.g., a novel dinosaur), or a novel picture from outside the category (e.g., a novel fish), infants looked more at the latter, indicating that they had formed a basic-level category. Importantly, this was only the case when a word (in a phrasal frame, like “a dino!”) was presented with the pictures. When a tone was presented instead, infants looked equally at the novel within-category exemplar as at the novel outside-of-the-category exemplar. Infants of older ages behave similarly in these tasks (Balaban & Waxman, 1997; Waxman & Booth, 2001). Moreover at 13 months of
age, other types of categorization effects were observed: infants could form superordinate categories (e.g., *animal*) in the presence of a word label, as opposed to no label, when infants were given real objects (Waxman & Markow, 1995).

At 9 and 10 months of age, infants also use words to track the appearance of *kinds* of things. In studies with this flavor, infants see one kind of object move out from one side of an occluder, and another kind of object emerge from the other side of the occluder. If 10-month-old infants never see two separate objects appear at the same time, they only expect one object to appear when the occluder is revealed, indicating that they were only able to use spatiotemporal information to individuate a single object. Twelve-month-old infants expected two objects, indicating that they could use perceptual features to track objects (Xu & Carey, 1996; Xu, Carey, & Quint, 2004). Crucially, even though 10-month-olds normally fail to detect changes in perceptual features, these infants are able to track and identify when distinct words, but not distinct emotional expressions or other kinds of attentional cues are paired with these objects (Xu, 2002). More evidence about infants' expectations of how *contrasting* words function come from Dewar & Xu (under review), who provided evidence that 9-month-olds expect there to be two kinds of objects present in a box when an experimenter looks into it (i.e., objects of different shapes, but not necessarily of different colors). Importantly, this was only the case when the experimenter mentioned two novel words (e.g., “Look, I see a fep! I see a dax!”). Infants expected two of the same kind of object when hearing only one novel word (e.g., “Look, I see a fep! I see a fep!”).

Moreover, words may also be used to guide expectations about deeper, non-visible properties of objects, like their functions. For example, both 9-month-old (Joshi &
Xu) and 13-month-old infants (Graham, Kilbreath, & Welder, 2004) are more likely to expect objects' functions to generalize to other objects with the same verbal label, somewhat independently of how perceptually similar those objects appear. In summary, these studies are taken as evidence that words may act as cues, or “placeholders,” for sortal-kind representations at this age (Xu, 2005).

Despite the controls used by Waxman, Xu, and colleagues, however, some evidence suggests that these effects are not limited to words, per se. For example, Fulkerson & Haaf (2003) reported that 9-month-olds seemed to form a basic-level conceptual category when a series of within-category toys were labeled with either a word or another type of sound (i.e., nonsense speech sounds like “sa sa sasasa,” or a tone-sequence). However, some specificity for words was observed; nine-month-olds were only able to create a superordinate category (i.e., animal) when given words, and not when given non-labeling sounds. Fulkerson & Haaf suggested that differences between their study and Balaban & Waxman (1997) may have been due to differences in methodology; they presented infants with real toys, while the former study used line drawings.

However other studies, with infants of slightly older ages, suggest that words are not unique cues for categorization, or signifiers of reference. Importantly, however, it seems that this is only the case when non-verbal cues are embedded in a clearly referential context. For example, Roberts and colleagues suggested that 15-month olds were able to use words and musical notes (when embedded in sentences) as cues to form superordinate categories (Roberts, 1995; Roberts & Jacob). Woodward and Hoyne (1999) suggested that 13-month olds could use word labels, as well as nonverbal sounds, to
retrieve an object when embedded in a sentence context (i.e., “Can you find the [BEEP],” where an electronic beep was used in place of a novel word). Campbell and Namy (2003) also reported that 13-month olds could accomplish this task with both words and nonverbal sounds, but again, only when that sound was delivered in a socio-referential context (i.e., only when the sounds came from the experimenter, and not from a baby monitor timed with experimenter interaction (see also Baldwin et al., 1996)). Thus non-speech cues may play a similar role that words do in both categorization and labeling, but importantly, only when these cues are embedded in the appropriate communicative context.

Evidence from older infants, 20 to 26 months of age, reinforces this view, suggesting that infants learn to select only those sounds which are important in language as they gain more experience with learning words. For example, twenty-month olds selectively learn to retrieve objects labeled by words, and ignore links between objects and non-linguistic sounds (Woodward & Hoyne, 1999). Similarly, Namy and Waxman (2002) reported that 26-, but not 18-month-olds selectively interpreted words, and not gestures, as labels.

Thus, there is a sizable body of evidence which suggests that from early in infancy, words act to influence the structure of concepts. There is, however, some debate as to whether “words” per se, are necessary to observe these effects. On one hand, words-but not tones, emotional expressions, or other non-speech sounds-seem to influence infants’ decisions about how to interpret various types of concepts in young infants. On the other hand, slightly older infants seem to be able to use some types of non-linguistic stimuli as referential labels, or cues for categorization. Finally, completing a U-shaped
function, infants of almost two years of age are selective about the signs which could possibly carry communicative function, even when embedded in socio-referential contexts. One possibility is, of course, that infants are aware of the special significance of words as communicative signs by at least 6 months, but they have no specific expectations about what purpose they serve until later in infancy, perhaps serving to categorize concepts. Around 9-10 months of age, infants may begin to realize that words play a role in establishing deeper conceptual categories like sortals or kinds, which make predictions about which perceptual features are important for categorization, and a further understanding of how to generalize non-visible, deep properties of concepts. By 12 months of age, infants begin to understand that words are actually signs, and that speech sounds serves to “stand for” or “map onto” referents. However, infants may also begin to extend semiotic function beyond simply words, and to any signifier which has communicative function (i.e., tones or beeps embedded in communicative contexts). Finally, as infants become toddlers, they learn that, by convention, linguistic signs are psychological links between speech sounds and concepts (Namy, Campbell, & Tomasello, 2004). Words are of a specific form, as agreed upon by a linguistic community, and kinds of signs that fall outside of communal use are ignored, including gestures and non-linguistic sounds.

2.1.3 Concepts-as-cues-for-words

The notion that the linguistic sounds are crucially linked to meaning is certainly an old one, pre-dating thought from either Saussure or linguists of the Prague school (i.e., Trubetskoy, Jakobson, others). However, it was these linguists who wrote prolifically about these ideas, defining a “phoneme,” which is a unit of sound defined only in relation
to other sounds—by its complementary distribution with other sounds among the words in a given language. Thus, any language has a unique phonemic inventory, consisting of sounds which all signify meaning. In English, for example, the phoneme /d/ distinguished “dark” from “ark,” or also from “bark.” Other types of non-phonemic speech sounds include allophonic variants of a phoneme, which do not signify meaningful differences: for example, the [d] in “this dog” versus the [D] in “our dog.” These sounds are articulated differently, but are both perceived by English speakers as one category (Werker & Tees, 1984b). Notably, these sounds actually have phonemic status in Hindi; put in another way, Hindi has a phonemic contrast for the word branch (i.e., /dal/) and lentil chutney (i.e., /Dal/), while English only has one word: doll (i.e., /dal/ in an upper-Midwestern accent).

Many linguists assumed that phonemes had psychological implications, particularly Trubetskoy, who made psychological claims about perception of sounds being mediated by native-language phonology. Saussure took a particularly strong stance, suggesting that there were absolutely no categories of sound without meaning (and vice versa). However, psychological and typological evidence suggests that both perception of sounds and meaning do have some a priori constraints; some phonetic contrasts, for example, are universally easier or universally more difficult to perceive due to articulatory or acoustic factors (Kuhl & Miller, 1978; Pisoni, 1979; Werker & Lalonde, 1988). These constraints, of course, are part of what justifies the definition of allophones; there are some phonetic distinctions which have commonalities across many languages, but only a subset of these distinctions is meaningful in any given language. Of interest here, however, is whether a weaker version of this relativist position can be defined: how
might concepts guide—but not necessarily determine—infants’ selective sensitivity for
words and speech sounds?

This question has been less well-researched in psychological circles than its
converse, particularly in development. In the domain of vision, however, there is some
evidence that higher-levels of conceptual structure (or at least category structure) can
affect lower-levels of visual perception. Goldstone (1995) reported that matching of
color-hues was influenced by whether or not the hue was presented in an alphabetic or
numeric shape. Hues that were presented in the same shape were judged to be closer
together. Similarly, linking arbitrary sortal concepts to one group of shapes (i.e.,
“salamander,” “dog,” or “recliner”) and arbitrary property concepts to another group of
objects (i.e., “slippery,” “furry,” or “comfortable”) affected performance on matching
tasks in accordance with semantic associations between the concepts and properties
(Gauthier, James, Curby, & Tarr, 2003). In other words, if two shapes were semantically
related, then it was easier to match those shapes.

Of particular interest, however, is the link between speech sounds and concepts.
What evidence is there that concepts can influence the way in which speech is perceived?
A few adult studies examined learning of speech sounds from feedback from higher-level
categories. Nusbaum and colleagues created a set of synthetic stimuli that were modeled
on the stop consonants /b/, /d/, and /g/ (Francis, Baldwin, & Nusbaum, 2000). Half of the
synthesized tokens had conflicting cues for place-of-articulation. For example, a
“conflicting” /b/ might have a formant transition for /b/, but a burst frequency for /d/.
Adults were trained to identify the set of synthesized speech sounds, both consistent and
conflicting (i.e., the conflicting sound from the example above was trained either as /b/ or
as /d/). One group was trained to respond on the basis of the formant-transition cue; another group was trained on the burst cue. In a test phase assessing categorization without any feedback to category membership, adults made their responses on the basis of the trained cue.

Hayes-Harb (2004) also taught English-speaking adults a non-native speech contrast. Adults were trained to distinguish unaspirated voiceless and voiced velar stops (i.e., [ga] and [ka]), which constitute a non-native phonetic contrast, normally difficult for English speakers to discriminate. In one condition, adults were given a 13 minute training period, where a picture of a rat was presented with [ga], and a picture of a pot was presented with [ka]. Discrimination of the speech contrast was better in a testing period that followed this condition, than in a testing period following a condition where only one picture was used for both sounds.

Taken together, these studies suggest that knowing category membership alone is a sufficient cue to selectively reorganize attention to acoustic parameters in speech. In other words, conceptual categories can affect the way that sounds are perceived. What evidence is there that this link is productive for word-learning? This hypothesis has not been tested directly; however, there is some preliminary evidence that learning links between sounds and concepts can influence perceptual sensitivity for speech. First, however, some relevant aspects of infants' early perceptual abilities are reviewed.

2.1.4 Early Speech Perception

Throughout the first-year of life, infants' developing phonetic sensitivities are reorganized with linguistic input from the environment. By about 10-12 months of age, infants show decreased performance in discriminating non-native phonetic contrasts,
discriminating only native ones (Werker & Tees, 1984a). For example, English-learning infants can discriminate dental (i.e., [d]) and retroflex (i.e., [D]) allophones of the phoneme /d/, but by 10-12 months of age these allophonic variants are not discriminated. Hindi-learning infants, however, continue to discriminate this contrast. This process has been termed infants’ functional reorganization of phonetic contrasts, because it was hypothesized that the meaningful status of this contrast in the infants’ native language was somehow driving the discrimination patterns (Pegg & Werker, 1997; Werker & Pegg, 1992; Werker, 1995; Werker & Tees, 1999). However, this language-specific decline in perceptual sensitivity is evident before infants have lexicons large enough to learn phonetic contrasts by comparing minimally different word forms.

One decidedly non-conceptual alternative to the functional reorganization hypothesis was proposed instead (Maye, Werker, & Gerken, 2002; Maye et al., in press). This approach has been termed “distributional learning,” because it appeals to frequency distributions in the input as a way for infants to learn phonetic categories. That is, simply being exposed to an adequate amount of natural language input may provide evidence for a bimodal distribution of sounds (Werker et al., in press). This may lead infants to posit language-specific phonetic categories, which only later are linked to meaning (Werker & Curtin, 2005). Maye et al. (2002) modeled this kind of exposure in the laboratory, exposing 6 to 8-month-old infants to either a bimodal or unimodal frequency distribution on a phonetic continuum spanning a non-native phonetic contrast. Infants who were exposed to a bimodal training phase were better than infants exposed to a unimodal training phase at discriminating the phonetic contrast. Similarly, Maye et al. (in press)
suggested that distributional training of a non-native contrast along one phonetic dimension was successfully generalized to other phonetic contexts by infants.

However, several reports have suggested that other kinds of cognitive abilities, besides distributional learning, influence phonetic discrimination (Kuhl, Tsao, & Liu, 2003; Lalonde & Werker, 1995). One line of evidence comes from a study where a non-native contrast was relearned at 9 months of age. Kuhl et al. (2003) reported that a Mandarin contrast was relearned only when infants were exposed to Mandarin in a socio-referential environment: that is, only in an environment where a human was interacting with the infant, and not when a videotaped or audio-taped version of the same human interacting with another infant was presented.

A tempting conclusion to draw is one that suggests that only in the socio-referential environment can infants successfully be taught to make links between objects and speech sounds. This concept-sound link essentially “re-teaches” a functional contrast between two sounds that Mandarin maintains, but English collapses into a single category. However, Kuhl et al. (2003) only measured the overall effect of social environment on discrimination performance, and could not precisely manipulate all of the variables between social and non-social conditions. Related to this point, Yoshida, Pons, & Werker (in prep) suggest that attention, which was not controlled in the Kuhl et al. study, affects learning from distributional input, even at 10 months of age. In other words, if infants at 9 months of age in Kuhl et al. (2003) paid more attention in the context of real interaction, they may still have relied on a distributional learning strategy, only having processed more in a more attentive condition.
Thus, there is some preliminary evidence that some factors beyond simply the statistical property of the input influences perceptual sensitivity. One possibility is that attention modulates learning, while a second non-mutually exclusive possibility is that higher-level conceptual learning can also guide phonetic sensitivity in human infants and adults. The question of interest, then, is what drives the process? Could it be related to the kind of phenomena noted in the words-as-cues-for-concepts section above; might hearing sounds in the presence of possible referents play a role in changing infants’ discrimination patterns?

2.2 Study 1

Whether or not infants know that concepts should be linked to different word forms has not been explicitly tested with young infants. This notion of “different words” takes advantage of categorization studies in the domain of speech, specifically learning of phonetic categories (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984a). While previous studies have suggested that infants as young as 6-8 months of age can learn phonetic categories from distributional characteristics of the input, this present study explores another hypothesis. While infants at this age do not have a lexicon large enough to learn phonemes from comparing minimal pairs, another possibility is that hearing words in the presence of objects is enough to drive perceptual categorization. Similar to the way that conceptual categories are influenced by the mere co-occurrence of words, is the categorization of minimally different word forms similarly influenced by the presence of labelable objects?

To answer this question, a study modeled after existing phonetic training studies was carried out. Because both statistical characteristics of the input, as well as attentional
factors are known to influence the effectiveness of training, an attempt was made to control these variables. In Study 1, infants 9 months of age participated with their parents in a looking-time study. Infants this age show declining discrimination ability for some non-native phonetic contrasts (Werker & Tees, 1984a), but can also improve discrimination ability if given particular types of familiarization before perceptual sensitivity is tested (Kuhl et al., 2003; Yoshida et al., in prep). Furthermore, performance in cognitive tasks with infants this age is influenced by the presence of a word, suggesting a link between concepts and words (Xu, 2002; Balaban & Waxman, 1997).

Infants were exposed to a training phase, and then a test phase assessing phonetic discrimination. Infants were trained to discriminate a non-native phonetic contrast by exposing them to naturally produced exemplars from each non-native category. A Hindi dental-retroflex contrast was used; infants were trained to discriminate syllables of the form \([\text{da}]\) from those of the form \([\text{Da}]\), a contrast which previous research has suggested is normally difficult for English-learning infants of this age (Werker & Tees, 1984). Crucially, this type training relied on the presence of novel, but distinct objects on a screen to signify a contrast between the two word forms. In this study, the two novel objects were shown on a screen during the training phase, and exemplars from one type of non-native phonetic category always occurred *concordantly* with one of the objects. In the test phase, discrimination was assessed in a similar way to previous studies, using an alternating/non-alternating paradigm to assess discrimination (Maye et al., 2002).

2.2.1 Method

In the final sample, 20 nine-month-old infants (10 female; mean age = 9;4 ; range = 8;13 - 9;19) were recruited from a database of parents who had previously expressed
interest in research studies. Care-givers and their infants arrived at our research centre, the study was explained, and finally a souvenir t-shirt and certificate were given as a token of thanks. Infants were exposed to at least 80% English, and less than 1% of any South Asian language, which commonly have dental-retroflex distinctions, as measured by parental report. Data from an additional 7 infants were not included in the final analysis due to failure to meet language criteria (1 female), fussiness (3 males; 1 female), and experimenter error (2 males).

In a looking-time procedure, infants were first familiarized to two pairings of objects and speech sounds (familiarization phase), and then were tested immediately afterwards on discrimination abilities for the familiarized speech sounds (test phase). Total testing time was approximately 7 minutes.

Stimuli. To test categorization of non-native speech sounds, naturally produced infant-directed tokens were elicited from a native Hindi-speaking female (age = 35 years). Stimuli were recorded in a sound-attenuated booth on a Radio Shack unidirectional dynamic microphone (model 33-3009) and an AudioBuddy preamp set at maximum gain. Tokens consisting of CV-syllables (dental /da/ and retroflex /da/) were elicited in sentential frames, but were excised from the frames using commercially available sound-editing software (Adobe Audition, version 1.5). Six dental and six retroflex tokens which had a similar low-high pitch contours were selected. Four tokens of each kind were used in the familiarization phase, while the remaining two tokens of each kind were used in the test phase. Average length of the familiarization tokens was 527ms (dental = 523ms; retroflex = 531ms), and the average length of the test tokens was 503ms (dental = 506ms; retroflex = 500ms). Figure 1 plots the first 500ms of the formant
contours for F1, F2, and F3 of each token (formant contours generated from Praat, version 4.4.22).

**Figure 1:** Formant contours plotted for each individual token /da/

![Formant Contours for Speech Tokens](image)

**Procedure.** Infants were tested in a quiet, softly lit room while sitting in their care-giver’s lap. Care-givers were instructed not to speak, and to do their best to keep their infants calm while listening to masking music over headphones during the entire procedure; they were seated 36 in. away from a black curtain. A 42” plasma screen was positioned in the middle of the curtain, and a slit for a video camera (Sony Digicam) was positioned 22 in from the floor, and 6 in from the bottom edge of the screen. Stimulus presentation was controlled by computer software (Habit X: Leslie Cohen at the University of Texas, Austin), and sounds were presented free-field over speakers hidden behind the curtain at approximately 60-62dB.
The study began with a 12 second pre-test trial showing an attractive toy on the screen paired with a nonsense word. The infant-controlled familiarization phase began next, as one object of two different objects appeared on the left side of the screen. It paused for 1000ms, rotated as it moved to the right side of the screen (250ms), paused for another 1000ms, subsequently rotated as it moved back to the left side of the screen (250ms), and paused for another 1000ms before disappearing. If infants were still looking at the screen, the sequence was repeated, but if infants were not looking, then a colorful pattern appeared on the screen without any accompanying sound until the infant looked back to the screen. Furthermore, if infants did not look for more than 2 seconds while an object was on the screen, that sequence of object movement was repeated.

One of the four CV-exemplars from one non-native category was presented synchronously with movement of the object. Movement of the object and sound presentation were synchronized in order to increase the likelihood that infants would encode the sound-object pairing at this age (Gogate, Walker-Andrews, & Bahrick, 2001). The ISI of sound presentation varied between 1000ms and 1700ms, and eight sound stimuli (four unique tokens) were presented to infants with one object before the second object appeared on the screen. When a new object appeared on the screen, the CV-exemplars from the other non-native category were presented in a similar fashion. Object-sound pairings alternated in this way until infants accumulated 2.5 minutes of looking. Thus, in this familiarization phase, one object was always paired with four exemplars of one type of sound (see Illustration 1 for a schematic of one possible pairing). The length of the familiarization period lasted ~4-5 minutes, including delays introduced by the computer when loading stimuli, and delays associated with the time it
took for infants to re-fixate on the screen after becoming distracted. The pairing between dental and retroflex sounds and the two novel objects was counter-balanced across infants, as well as the order of dental- and retroflex- blocks.

**Illustration 1:** One possible pairing of objects and sounds, where individual exemplars are denoted by use of subscripts (dental = black font; retroflex = gray font)

\[
[da_1] [da_2] [da_3] [da_4]
\]

\[
[Da_1] [Da_2] [Da_3] [Da_4]
\]

As soon as infants had accumulated enough looking-time, the object on the display disappeared from the screen, and the test phase began. Because both categories of sounds were presented during the familiarization phase, a standard habituation-discrimination or novelty-preference paradigm was deemed to be unfeasible. Instead, an alternating/non-alternating paradigm was used to assess discrimination (Best & Jones, 1998; Maye et al., 2002). In this study a non-object (i.e., a checkerboard) was displayed visually, and infants were assumed to discriminate sound categories looked longer in
trials where separate tokens from the same sound category were presented (non-alternating trials) than in trials where separate tokens from different sound categories were presented (alternating trials). Infants who looked equally at the two types of trials were assumed not to discriminate these sound categories.

In the test phase, a version of the alternating/non-alternating discrimination paradigm that closely followed that of Maye et al. (2002) was presented to infants. Black-and-white checkerboards appeared on the screen in this test phase while looking-time was recorded. Each checkerboard trial lasted 10 seconds, and nine trials were presented in total. The 1st checkerboard was presented without any auditory stimuli to give infants a chance to look at the novel visual stimulus, since they had just completed the familiarization phase. When a trial finished, the same colorful pattern used to attract infants' attention in the familiarization phase was displayed if infants were no longer looking at the screen; otherwise, the next trial began immediately.

Eight more checkerboard trials were then presented with sound stimuli. As described above, four novel tokens (two dental; two retroflex) were used. Two types of trials were presented: non-alternating and alternating types. Non-alternating types of trials contained two tokens from within a phonetic category: either the two dental tokens, or the two retroflex tokens. Alternating types of trials contained two tokens, one from each of the two phonetic categories; all four possible combinations of two tokens were used for the alternating trials. In a single trial, sounds were presented with an ISI of one second, also with an SOA of about 1500ms (since all stimuli were approximately 500ms), and alternated between the two tokens for a total of 10 seconds. For all infants, the first checkerboard was silent, but for half the infants, the 2nd, 4th, 6th, and 8th trials were the
alternating type, and the 3rd, 5th, 7th, and 9th trials were the non-alternating type. For the other half of infants, the order of trial types was counter-balanced. Furthermore, all possible orders (dental non-alternating; retroflex non-alternating; four types of alternating trials) were presented across all infants in a maximally counter-balanced way.

**Analysis.** Videos of the test trials were digitized from DAT recordings using a Mac G5 running Final Cut Pro and converted to QuickTime movies. Looking-time to test trials containing auditory stimuli was coded frame-by-frame at a rate of 29.97 frames per second by a trained coder using customized computer scripts. Sounds from the 2nd and 3rd test trials were pooled so that each infant contributed to the means for both alternating and non-alternating trial types within the first pair of test trials containing auditory stimuli. For each infant, total looking in a 10-second trial was entered into a data table as Pair 1 with both alternating and non-alternating types, and a similar analysis was done for the 4th and 5th test trial (Pair 2), the 6th and 7th test trial (Pair 3), and the 8th and 9th test trial (Pair 4). Data from the table were entered into a 2 x 4 ANOVA with factors Type (alternating vs. non-alternating) x Pair (1-4). If infants discriminated the sound categories, it was expected that there would be no interaction, but a main effect of Type (i.e., non-alternating > alternating), and a main effect of Pair (i.e., infants showing a linear decrease in looking over the pairs of trials in the test phase). If infants did not discriminate the sound categories, it was expected that there would be no interaction or main effect of Type, but a main effect of Pair, as infants continued to look less over the test trials.
2.2.2 Results

Infants' looking time to eight trials in the test phase which contained sound stimuli was entered into an ANOVA with factors of Type and Pair. If infants could discriminate this non-native contrast, then it was predicted that they would look longer to one type of trial in this phase. Previous research suggests that infants usually look longer to the non-alternating trial type in this paradigm if they are able to discriminate the sounds. The pattern of results in this current study suggested that infants could indeed discriminate the sound categories after being exposed to the familiarization phase.

The 2 x 4 (Type x Pair) ANOVA yielded no significant interaction (p = 0.610). However, there was a main effect of Type (F(1,19) = 4.628; p < 0.05), indicating that infants looked on average longer at non-alternating (5.10s) compared to alternating (4.54s) trials in the test phase. Furthermore, there was also a main effect of Pair (F(3,57) = 4.937; p < 0.01), indicating that infants looked progressively less at each pair of test trials. A significant linear contrast (F(1,19) = 9.771, p < 0.01) was observed for this main effect. Looking time is charted in Figure 2.

Thus, these results suggest that infants were habituating to the stimuli in the test phase, but overall looked longer at the non-alternating trial type. The pattern of non-alternating trials receiving significantly more looking than alternating trials replicates previous patterns (Maye et al., 2002), and suggests that infants were able to discriminate the non-native contrast.
2.2.3 Discussion

Previous work suggests that infants 9 months of age are in the process of forming language-specific phonetic categories, and have some trouble discriminating non-native phonetic contrasts (Werker & Tees, 1984a). Yet, these results suggest that infants this age can recover discrimination ability for these contrasts (see also Kuhl et al., 2003). Specifically, infants might be able to use the co-occurrence of distinct visual stimuli with non-native speech sounds to classify phonetic information, and form perceptual categories. However, whether infants depended on the link between speech sounds and objects to form phonetic categories, rather simply relying on statistical cues also present in the familiarization phase, is still unclear. For example, because infants heard naturally produced exemplars from two phonetic categories, they were essentially exposed to a
bimodal cluster of sounds in phonetic space\textsuperscript{9}, similar to bimodal distributions used in previous studies on distributional learning. If infants did not pay attention to the visual stimuli at all, then the bimodal nature of the training phase, from a purely acoustic point-of-view, may still have been enough to allow infants to learn to discriminate these sounds. To discount this possibility, a second study was run. If infants were paying attention to the co-occurrence of the sounds and objects, then disrupting this co-occurrence (by presenting discordant pairings between sounds and objects) should also affect infants’ perceptual sensitivities.

2.3 Study 2

The rationale for this current study is based on the possibility that infants were not linking the speech sounds and visual stimuli in Study 1, but were depending instead on the distribution of sounds in the familiarization phase to learn the phonetic contrast. If this latter hypothesis were true, then presenting the infants with the same auditory input, but disrupting the concordance between pairings of sounds and objects should not affect infants’ ability to discriminate the sounds in the test phase. However, if the pairing of object and sound was driving perceptual reorganization, then presenting infants with a discordant familiarization phase should similarly disrupt discrimination in the test phase. In Study 2, a familiarization phase which consisted of the same auditory and visual stimuli as Study 1 was used, but presented in a discordant fashion; sometimes dental /da/ tokens would be paired with one object, but sometimes they would also be paired with the second object, and vice versa for the retroflex /Da/ tokens (see Illustration 2). Importantly, the distribution of the sounds or objects within a single modality was similar to that in Study 1.
Illustration 2: Sample of Pairings used in Study 2

2.3.1 Method

In this sample, 20 nine-month-old infants (10 female; mean age = 9;4; range = 8;14 - 9;22) were recruited from a database of parents who had previously expressed interest in research studies. In all respects, the procedure was identical to Study 1 with the exception of the following differences in the familiarization phase. Data from an additional 5 infants were not included in the final analysis due to fussiness (1 male; 2 females), experimenter error (1 female), and being out of range of the camera for a prolonged period of time (1 female).

In Study 1, each infant was exposed to two unique types of trials: one of two sounds was always exclusively paired with one of two objects. A trial consisted of an object presented on a screen with eight exemplars of naturally produced speech tokens.
from one non-native phonetic category (4 unique tokens). Here, trials were the same as those used in Study 1: eight exemplars (4 unique tokens) from one phonetic category were paired with an object. However, four unique types of trials were presented to infants in the familiarization phase: the full permutation of pairings between two objects and two sound types. Thus, infants received the same distribution of sounds and objects, but only the pairings between the types of sounds and objects were discordant. As in Study 1, an infant-controlled familiarization was used, and a total of 2.5 minutes of looking was accumulated before moving onto the test phase. The number of trials that were played in the familiarization phase, which differed for each infant due to differences in looking patterns, did not differ from Study 1 (Mean = 56.3 trials in Study 1 vs. mean = 56.8 trials in Study 2, p = 0.628). The test phase was identical to Study 1.

**Analysis.** The analysis of the data in Study 2 was identical to that in Study 1. However, the predictions differed. In this study, unlike in the previous one, infants were not expected to be successful at discriminating the non-native contrast in the test phase if their perceptual sensitivities were influenced by links between sounds and the objects. If this was indeed the case, no interaction or main effect involving Trial Type was expected. However, if infants were not influenced by this pairing, then the results should parallel those in the previous study, showing a main effect of Trial Type. Infants were also expected to habituate to the test trials over the whole testing phase, and a main effect of Trial Pair was expected.
2.3.2 Results

The pattern of results suggested that infants were not successful in discriminating the sound categories after being exposed to the modified, discordant familiarization phase. Ten of twenty infants looked longer to the non-alternating trials, in contrast to Study 1, where fourteen of twenty infants looked longer to the non-alternating trials. The 2 x 4 (Type x Pair) ANOVA yielded no significant interaction (p = 0.501). Nor was there a main of effect of Type (p = 0.690), indicating that infants looked equally at non-alternating (5.13s) compared to alternating (4.94s) trials in the test phase. However, there was a main effect of Pair (F(3,60) = 12.382; p < 0.001), indicating that infants looked progressively less at each pair of test trials. A significant linear (F(1,20) = 19.870, p < 0.001) and quadratic (F(1,20) = 6.864, p < 0.05) contrast was observed for this main effect. Thus these results suggest that infants were habituating to the stimuli in the test phase, but overall looked equally at the non-alternating and alternating trial types. In summary, and unlike in Study 1, this group of infants was not able to discriminate the non-native contrast. Looking time is charted in Figure 3.
Because infants’ declining looking over the duration of the test phase did not play an important theoretical role in these studies, further analysis collapsed over all trial pairs from Study 1 and Study 2 producing overall means for non-alternating and alternating trial types. Thus, a 2 x 2 between-within ANOVA was conducted with a between-subject factor of Study (1 or 2), and a within-subject factor of Trial Type (non-alternating vs. alternating). However, no significant effects were found. This lack of an effect may have been due to the relatively weak effects found in Study 1. Figure 4 plots this data.
2.3.3 Discussion

The hypothesis that infants depended on a distributional learning mechanism to succeed in discriminating the non-native contrast in Study 1 is weakened by the null effect in Study 2. The distribution of auditory stimuli in both studies was similar, and moreover, attentional factors were controlled by familiarizing infants until their looking times reached a pre-set criterion. However, looking patterns still seemed to differ between the two studies. If infants were depending instead on the link between the object and speech sounds to categorize the phonetic information in the familiarization phase, then this pattern of results would be predicted. In Study 1, the concordant pairings were a reliable cue to phonetic category membership, but in Study 2, this cue was deliberately disrupted by using discordant pairings.
Although it seems clear that infants succeeded in discriminating the non-native contrast in Study 1 by relying on the pairing between sound and object, whether infants failed to discriminate the phonetic contrast in Study 2 because of the absence of concordant pairings, or because the discordant pairings were actively confusing for infants is unknown. In the case of discordance, evidence for two distinct categories was not given through the pairings, but nor were infants given explicit evidence that they should collapse both phonetic categories into a single one. Evidence for this latter possibility would be more explicit if only one object were present in the familiarization phase. However, in the interest of making Study 2 comparable to Study 1, both phonetic categories were paired with both distinct objects—not a situation where cues for single-category formation were unambiguous.

Yet, even when information from pairings between objects and sounds were not useful, infants were still not able to rely strictly on the bimodal nature of the sound patterns to tune their perceptual sensitivity. Comparing these methods to those used in previous studies of distributional learning (Maye et al., 2002), neither the length of the familiarization phase (2 minutes vs. 2.5 minutes), nor the rate of stimulus presentation (one stimuli every 1000ms vs. every 1000-1700ms here) differed by very much. This raises the interesting possibility that, when objects are present and may act as possible cues, taking the pairing between object and sound into account might be a mandatory process.

However, there are several reasons that this preliminary conclusion may be premature. First, simply having dynamic visual displays in the current study versus static visual displays used in other statistical learning and phonetic training studies may have
increased the overall difficulty of the task (Robinson & Sloutsky, 2004). This raises the possibility that only in Study 1, when visual cues correlated with phonetic categories, did infants have the extra information needed to succeed in an otherwise difficult task. In Study 2, in the absence of these cues, infants may still have failed to learn contrastive phonetic categories due to the inherent difficulty of this task, and without any specific additive processing load from trying to make sense of discordant pairings.

Furthermore, these studies differ in three important respects from other statistical learning studies, which may offer explanations as to why infants did not simply rely on the distribution of phonetic information in the familiarization phase to discriminate the sounds in the test phase. First, the present study used naturally produced tokens, rather than edited speech. Thus, there was more variability in the acoustic cues, perhaps making statistical generalization harder than with more controlled stimuli. Second, the formant values of these stimuli were not distributed along a continuum from a dental to retroflex place-of-articulation; rather they were clustered into two modal groups (see Figure 1). Generalizing from this kind of distribution might be harder than from one with more kurtosis. Third, there were only four unique tokens used in the training phase, while other studies of distributional learning have used continua with at least 8 members. Moreover, testing here was assessed with novel tokens (although from the same speaker, recorded in the same session), but this was not the case in other distributional learning studies (Maye et al., 2002; Maye et al., 2005).

To summarize, the precise reason why infants may have failed in Study 2 is unclear. Thus two possibilities are proposed. First, the absence of a consistent pairing between sound and object made an otherwise difficult task too hard for young infants to
succeed at. Second, infants might have linked sounds and objects in a mandatory fashion, and difficulty arose from infants’ processing of the essentially random pattern of pairings in Study 2. More research will be needed to distinguish between these two possibilities.

2.4 General Discussion

These studies show that 9-month-old English-learning infants can discriminate a non-native phonetic contrast after being exposed to 2.5 minutes of a familiarization phase where exemplars from each non-native category were paired with a distinct object. Performance was not successful when the same sounds and objects were presented, but the pairings were discordant. At a minimum, these results suggest that the co-occurrence of visual cues promotes the formation of distinct phonetic categories, while category formation is disrupted when these cues are not informative. Such a claim is significant in the following way: it provides one possible mechanism of functional reorganization, suggesting that infants’ perceptual sensitivities for speech may be re-organized in the first 6-12 months of life due, in part at least, to speech sounds’ co-occurrence with distinct visual patterns in the infants’ world. This would be conversely related to the way that the formation of some conceptual categories may be aided by pairing the a word with pictures of within-category exemplars (Balaban & Waxman, 1997). It is important to note, however, that this proposal only offers one possible mechanism for functional reorganization. Importantly, statistical approaches, the current strategy here, and any number of other possible strategies may conspire to guide infants’ formation of language-specific phonetic categories. The relative strength of each learning strategy may vary at different points in development, and further work will need to be conducted in order to distinguish between them.
Furthermore, in this first set of studies, movement of the objects on the screen and presentation of the sound were always synchronous. This decision has the disadvantage of potentially rendering the word forms as intermodal properties of the objects' movement, and is not the strongest case to argue for semiotic influences in the familiarization phases used here. However, as a first step towards arguing for perceptual learning from conceptual cues, it is important to establish that objects can potentially act as cues to categorize sounds, and previous research suggests that synchrony improves the likelihood that young infants will dishabituate to a link between a sound and an object (Gogate & Bahrick, 1998; Gogate & Bahrick, 2001). Thus, the decision to synchronize the sound and object was made in order to improve the likelihood that infants would succeed at this task. Furthermore, observational studies of word-learning have suggested that intersensory input is in fact common in mother-child interactions, particularly in ostensive labeling situations (Gogate, Bahrick, & Watson, 2000), and that it correlates with word-learning abilities (Gogate, Bolzani, & Betancourt, 2006). This evidence that infants are actually exposed to simultaneous visual and auditory cues only increases the plausibility that infants actually depend on making this cross-modal link when forming perceptual categories.

In summary, these results revive, in spirit at least, Werker and colleagues’ earlier notion of functional reorganization (Werker & Pegg, 1992; Werker, 1995; Werker & Tees, 1999). Specifically, it provides specific evidence for two theories of how these links between sounds and objects may be important in reorganizing perceptual sensitivity for speech. First, infants may be sensitive to correlated cross-modal cues (i.e., between two distinctive visual cues and two similar categories of speech sounds), and can use
these cues for perceptual categorization after only very brief exposure. Indeed, one might predict that infants receive input such that the appearance of similar visual stimuli is correlated with and predictive of the occurrence of similar speech sounds. Since Gogate et al. (2005; 2006) have provided evidence that infants receive input like this in certain contexts, preliminary support for this view is provided. However, without detailed empirical evidence that "strong-enough" correlations actually exist in the whole of infants' day-to-day perceptual experience, including contexts where infants are not taught words, one might caution against such an unconstrained interpretation. In particular, it might result in "promiscuous correlations" between any recurring visual stimulus and native-language categories which ought to be grouped separately.

Another hypothesis that these results support is one where infants 9 months of age use objects as conceptual cues which signify the presence of distinct phonetic categories. This possibility is reminiscent of the idea that sounds and concepts are linked, bi-directionally, in an infant-version of a semiotic sign. That is, just as words (or any signal embedded in a communicative context) seems to have some special role in the process of forming and identifying conceptual categories (Waxman, 2004; Xu, 2002), concepts may have a similar effect on speech sounds. For example, the visual stimuli used as cues for categorization were also novel and cohesively moving objects, possible referents for the syllables with which they were paired. Thus, these results provide tentative evidence that infants are able to use links between sounds and possible referents to categorize phonetic information. One interesting test of this hypothesis would be to test whether or not phonetic categorization would still be observed if the objects used in the familiarization phase were perceptually very distinct, but conceptually more similar. For
example, one prediction might be that objects which differed in *kind* would produce this effect, but objects that did not differ in kind (yet were perceptually still distinct), would not.

Developmentally, this link between concepts and sounds connects to existing literature in infants’ perception of words in interesting ways. Importantly, a clear division needs to be made between links of the sort described in the current study, and learning to remember actual *word forms*. For example, some evidence suggests that young infants begin to *listen* for meaning in possible word forms at a young age (Werker & Pegg, 1992; Werker & Tees, 1999). For example, infants as young as 8 months of age selectively prefer passages containing words that they were previously familiarized with, but not passages containing slight mispronunciations (Jusczyk & Aslin, 1995). Furthermore, infants as young as 10 and 11 months of age show different patterns of looking to lists of common as opposed to rare words (Halle & Boysson-Bardies, 1994).

Yet, even though the current study suggests that infants categorize word forms in the presence of objects, their knowledge of these phonetic categories may not be used to *remember* word forms accurately. For example, when mispronunciations occur in non-initial (e.g., *paart* vs. *paarp*) or non-stressed positions (e.g., *dinner* vs. *ninner*), 10- and 11-month olds treat these mispronunciations like as if they were frequent word forms ((Halle, Segui, Frauenfelder, & Meunier, 1998; Swingley, 2005; Vihman, Nakai, DePaolis, & Halle, 2004). Moreover, even if infants show perceptual sensitivity for phonetic contrasts used to distinguish different word forms, learning to remember the specific links between these minimally different word forms and different objects is still quite difficult (Stager & Werker, 1997). In summary, these results do not necessarily
show that learning words or even of word forms, as such, is what drives tuning of perceptual sensitivity. Rather, early expectations about a relation between speech sounds and distinct concepts may influence the categorization of phonetic information without positing a reference relationship, per se, or memory for a specific associative link.

In conclusion, these results suggest that semiotic principles may inform developmental research in novel ways, suggesting new directions for research, and providing ways to understand existing evidence. Current empirical work provides preliminary evidence that young infants have an emerging understanding of signs which may include a loose understanding of a link between a word form and an object, but which does encompass full reference. This link may allow a bi-directional link between sounds and concepts, forming the basis for both conceptual and phonetic categorization in infancy. And while the current evidence does not unequivocally support the idea that concepts will influence phonetic perception by positing distinct word forms, at a minimum these results suggest that infants can use visual cues as a cue for phonetic categorization. This provides support for the idea that infants may depend on language-specific input from both speech and other extra-modal cues to functionally reorganize their perceptual sensitivity.
Notes

Such divisions can be traced, however, back to both Locke and Aristotle; the former making distinctions between words, things, & ideas in his Essay Concerning Human Understanding (1690) and the latter making similar distinctions in Poetics (~350 BCE).

These departments, today, would be called, “Indo-European.”

This name is actually pronounced “PURSE.”

The case of Japanese mimetics seems to be controversial as to whether or not it constitutes a “special” case of language use (Hamano, 1998).

This use of “symbol” differs from both Saussure and DeLoache’s rather confusing use of the same word (DeLoache, 2005; Saussure, 1983). Saussure’s use of “symbol” corresponded more with Peirce’s use of the word, “index,” while DeLoache’s use of the same word corresponds more appropriately semioticians’ use of the word “sign,” meaning any entity which exists independently from that which it denotes (DeLoache, 2005; Sebeok, 1991). To add to the confusion, her work mostly deals with icons, like pictures, maps, and scale-models, and not symbols at all, as Peirce and the field of semiotics tend to define that word. However, in following work from semiotics, I propose from to stick with use of the four terms as described here: sign (the super-ordinate term for the following three), icon, index, and symbol.

References omitted for formatting purposes; see text for more complete description.

See also Xu & Baker, 2005 for a related task.

Allophones are stable phonetic instantiations of a sound, which may or may not have phonemic status in another language. They are usually annotated with brackets (e.g., [d]), while phonemes are usually annotated with slashes (e.g., /d/)

Natural tokens are notoriously hard to place on a continuum, particularly in the case of place-of-articulation differences, where several acoustic parameters simultaneously change. To investigate questions of distributions, synthesized or edited natural speech are commonly used. However, in this first study, naturally elicited speech was used to maximize the ecological validity of this kind of manipulation.

Moreover, theses tokens were purposefully elicited in pre-determined naming phrases, such that normal variability from co-articulated natural speech would be controlled.

Dutch for “horse,” as this study was carried out in the Netherlands.
Bibliography


Dewar, K., & Xu, F. (under review). Do 9-month-old infants expect distinct words to refer to kinds?


Joshi, A., & Xu, F. *Inductive inference, artifact kind concepts, and language*. Unpublished manuscript.


Appendix A

Copy of UBC Research Ethics Board's Certificate of Approval