1-1-2013

Word Learning in 6-16 Month Old Infants

Elika Bergelson

University of Pennsylvania, elika.bergelson@gmail.com

Follow this and additional works at: http://repository.upenn.edu/edissertations

Part of the Cognitive Psychology Commons, and the Developmental Psychology Commons

Recommended Citation

Bergelson, Elika, "Word Learning in 6-16 Month Old Infants" (2013). Publicly Accessible Penn Dissertations. 735.
http://repository.upenn.edu/edissertations/735

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/735
For more information, please contact libraryrepository@pobox.upenn.edu.
Word Learning in 6-16 Month Old Infants

Abstract
Understanding words requires infants to not only isolate words from the speech around them and delineate concepts from their world experience, but also to establish which words signify which concepts, in all and only the right set of circumstances. Previous research places the onset of this ability around infants’ first birthdays, at which point they have begun to solidify their native language phonology, and have learned a good deal about categories, objects, and people. In this dissertation, I present research that alters this accepted timeline. In Study 1, I find that by 6 months of age, infants demonstrate understanding of around a dozen words for foods and body parts. Around 13-14 months of age, performance increases significantly. In Study 2, I find that for a set of early non-nouns, e.g. `uh-oh' and `eat', infants do not show understanding until 10 months, but again show a big comprehension boost around 13-14 months. I discuss possible reasons for the onset of noun-comprehension at 6 months, the relative delay in non-noun comprehension, and the performance boost for both word-types around 13-14 months. In Study 3, I replicate and extend Study 1’s findings, showing that around 6 months infants also understand food and body-part words when these words are spoken by a new person, but conversely, by 12 months, show poor word comprehension if a single vowel in the word is changed, even when the speaker is highly familiar. Taken together, these results suggest that word learning begins before infants have fully solidified their native language phonology, that certain generalizations about words are available to infants at the outset of word comprehension, and that infants are able to learn words for complex object and event categories before their first birthday. Implications for language acquisition and cognitive development more broadly are discussed.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Psychology

First Advisor
Daniel Swingley

Keywords
cognitive development, infancy, language acquisition, word learning

Subject Categories
Cognitive Psychology | Developmental Psychology | Psychology

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/735
WORD LEARNING IN 6-16 MONTH OLD INFANTS
Elika Bergelson

A DISSERTATION

in

Psychology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2013

Supervisor of Dissertation

Daniel Swingley

Associate Professor of Psychology

Graduate Group Chairperson

John Trueswell, Professor of Psychology

Dissertation Committee

Delphine Dahan, Associate Professor of Psychology

Daniel Swingley, Associate Professor of Psychology

John Trueswell, Professor of Psychology
WORD LEARNING IN 6-16 MONTH OLD INFANTS

COPYRIGHT

2013

Erika Bergelson

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License

To view a copy of this license, visit

http://creativecommons.org/licenses/by-ny-sa/2.0/
ACKNOWLEDGMENT

This dissertation would never have existed were it not for the army of advisors, mentors, colleagues, friends, and family who have supported me, in every way, for the past five years. First and foremost, I’d like to thank my advisor, Dan; I could not have picked a better graduate advisor. Without your patient help, prodding questions, excellent advice, high standards, comforting support, confidence that I could figure things out, pickiness about aesthetics and vocabulary, and genuine caring mentorship and friendship, I could never have accomplished the research in this dissertation. Thanks to John for your availability and advice, and Delphine for your mentorship, collaboration, and friendship. Thanks Lila for your sage mentorship, willing advice, support, excitement, and never-ending sushi. Thanks to Robert for your interest and input, from my 699, to my major areas, to my post-grad life.

I also grew greatly from the wisdom and friendship of my psych and IRCS cohort mates. Thanks to Andy, for always offering advice on any issue, from the little ones to the really big ones, to Joe for endless statistical patience and R wisdom, to Ann for food and fun, to Carolyn for putting up with my endless barrage of question, to Noah for being my academic big brother, and to Constantine, Emily, Isaac, Chaz, Emlen, Suzanne, Dahlia, and Pavel for being my pals.

Next my UMD and NYU support teams needs thanked, for continuing advice, and a good dose of linguistic grounding: thanks to Jeff, Bill, Colin, Norbert, David, Brian, and Annie. Thanks to Dave for pointing out my idiolect and for offering moral support through some tumultuous years. Thanks to Cousin Gary for getting me going down this path. Thanks too to my conference buddies and future colleagues: Katherine, Krista, Henny, Casey, Adriana, and Celeste.

Without a grounded home life, I could never have maintained by sanity at work: thanks to Rose (copyeditor extraordinaire) and Emily for 4 years of the best room-mate BFFs I’ve ever had, and to Casey, Ari, Paul, Dan, Danielle, and Lydia for being my pals. Thanks to Zach for being my rock these past two crazy years, and for your endless patience, love, support, good listening, and hugs. Thanks to my multi-lingual parents and siblings who got me interested in language in the first place, for your interest in my work and your support of me, in every way, throughout the years.

I appreciate you all in a way that the words on this page will never express. Thank you.
ABSTRACT
WORD LEARNING IN 6-16 MONTH OLD INFANTS
Elika Bergelson

Daniel Swingley

Understanding words requires infants to not only isolate words from the speech around them and delineate concepts from their world experience, but also to establish which words signify which concepts, in all and only the right set of circumstances. Previous research places the onset of this ability around infants’ first birthdays, at which point they have begun to solidify their native language phonology, and have learned a good deal about categories, objects, and people. In this dissertation, I present research that alters this accepted timeline. In Study 1, I find that by 6 months of age, infants demonstrate understanding of around a dozen words for foods and body parts. Around 13-14 months of age, performance increases significantly. In Study 2, I find that for a set of early non-nouns, e.g. ‘uh-oh’ and ‘eat’, infants do not show understanding until 10 months, but again show a big comprehension boost around 13-14 months. I discuss possible reasons for the onset of noun-comprehension at 6 months, the relative delay in non-noun comprehension, and the performance boost for both word-types around 13-14 months. In Study 3, I replicate and extend Study 1’s findings, showing that around 6 months infants also understand food and body-part words when these words are spoken by a new person, but conversely, by 12 months, show poor word comprehension if a single vowel in the word is changed, even when the speaker is highly familiar. Taken together, these results suggest that word learning begins before infants have fully solidified their native language phonology, that certain generalizations about words are available to infants at the outset of word comprehension, and that infants are able to learn words for complex object and event categories before their first birthday. Implications for language acquisition and cognitive development more broadly are discussed.
TABLE OF CONTENTS

ABSTRACT......................................................................................................................... iv
LIST OF TABLES .................................................................................................................. vi
LIST OF FIGURES ............................................................................................................... vii
Chapter 1: Introduction ...................................................................................................... 1
   The Word-form Problem ............................................................................................... 2
   The Concept Problem ................................................................................................. 14
   The Mapping Problem ............................................................................................... 22
   Word Comprehension Theories and Summary ......................................................... 31
Chapter 2: At Six to Nine Months, Human Infants Know the ........................................ 34
   Meanings of Many Common Nouns ........................................................................... 34
Chapter 3: The Acquisition of Abstract Words by Young Infants .................................... 49
Chapter 4: Young Infants’ Generalizations over Word-Forms during
   Comprehension ........................................................................................................... 59
Chapter 5: Conclusions, Further Directions, and Discussion of Word-Learning
   Mechanisms ................................................................................................................ 80
      Summary of Dissertation ......................................................................................... 80
      Birth to 6 months .................................................................................................... 81
      Between 6-9 and 10-13 months ............................................................................ 85
      Growing Phonetic Precision ................................................................................... 87
      Improvement at 13-14 months ............................................................................... 88
      Cautionary Notes .................................................................................................... 98
BIBLIOGRAPHY ................................................................................................................ 101
   Chapter 1 References: .............................................................................................. 101
   Chapter 2 References: .............................................................................................. 109
   Chapter 3 References: .............................................................................................. 111
   Chapter 4 References: .............................................................................................. 113
LIST OF TABLES

Chapter 2, Table 1: Target items .................................................................36
Chapter 2, Table 2: Summary statistics for CDI ........................................43
Chapter 2, Table S1: Model Details .............................................................48
Chapter 4, Table 1: Stimuli used in Experiment 2 ........................................71
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Figure Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 2</td>
<td>Figure 1: Experimental setup</td>
<td>36</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>Figure 2: Subject and item/pair means</td>
<td>40</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>Figure 3: Timecourse of infants’ picture fixation</td>
<td>40</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>Figure S1: Sample visual stimuli and regions of interest</td>
<td>47</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>Figure S2: Experimental timeline</td>
<td>47</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Figure 1: Sample stills from video stimuli</td>
<td>52</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Figure 2: Target looking performance in each infant</td>
<td>54</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Figure 3: Item-pair mean difference scores</td>
<td>54</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Figure 1: Experimental setup</td>
<td>64</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Figure 2: Subject means for B&amp;S12 and experiment 1 and 2</td>
<td>66</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

“Words could never be more than hints to what is really an enormously complex process, one which we scarcely understand even vaguely.” (Macnamara, 1982, p. 98)
“Nobody knows how children learn the meanings of words.” (P. Bloom, 2002, p. 262)

In this dissertation, I examine infants’ early word learning. While the received wisdom is that infants do not learn words until they are around one year of age (P. Bloom, 2002; Fisher & Gleitman, 2002; Kuhl, 2011; Macnamara, 1982), I will present evidence that this assumed timeline is wrong. However, a change in timeline is hardly worth noting in and of itself: for many results in cognitive development, with the right methodology, researchers can show that infants perform in a clever way “earlier than we thought.” What is worth noting are the repercussions of a shifted timeline of early word learning on the rest of language acquisition, and more broadly, on our understanding of infants’ cognitive and social development.

Many tomes have been written about early word learning, and while researchers differ in how “hard” they think the problem is, they generally agree that the learner is faced with a three-part problem: finding the chunks of the auditory stream that stand alone as words (Word-form Problem), identifying the aspects of the world that are individuated as lexicalized concepts (Concept Problem), and figuring out which words signify which concepts within their linguistic community (Mapping Problem). But which words do they learn first, and why? What abilities must they have, at a minimum, to succeed in this task, and what abilities are not needed? Do infants make the right assumptions about what kinds of properties matter for word meaning, and what kinds do not? In the chapters that follow, I suggest answers to these questions.

In the work that follows, I chose to focus on 6-16 month olds. The lower end of this age range was selected because this was the earliest age infants had been shown to understand any words, namely, ‘mommy’ and ‘daddy’ (Tincoff & Jusczyk, 1999). While proper nouns (which includes ‘mommy’ and ‘daddy’ for infants) have many properties that differentiate them from common nouns (Macnamara, 1982, Chapter 2), this seemed like a reasonable lower bound to search within for word comprehension. That is, at this age, it was likely we would find little evidence of word meaning knowledge. Previous work on common nouns did not demonstrate understanding in the laboratory until 13 months of age (Thomas, Campos, Shucard, Ramsay, & Shucard, 1981), though diary studies suggest earlier comprehension (Bates, 1993; Benedict, 1979; Dale & Fenson, 1996). The older end (16 months) was selected as a ceiling: in all the studies described in this dissertation, we found that performance by infants older than around 14 months was very strong, so there was no need to extend the range further (though in Chapter 2, we did test an 18-20 month group). Moreover, we were interested in the very early beginnings of word-learning: beyond 16 months infants have tools in the word-learning toolbox that younger infants do not, such as the preliminaries of syntax, the ability to use mutual exclusivity, a production vocabulary, etc. (Gleitman, 1990; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Jin & Fisher, 2013). Thus, the main age of interest in the work
below is from 6-12 months, with older groups of infants included for the sake of comparison.

I now turn to a description of the three-part word-learning problem described above, with a focus on the first year of life.

**The Word-form Problem**

In the first 16 months, infants perform impressive feats of learning in their native language. In comprehension, the main focus of this dissertation, they learn about linguistic properties across varying levels of representation. In the sound structure domain, infants learn to discriminate various aspects of the speech stream, deduce their language-specific phonemes, segment words out of the speech stream they hear, store word-forms with a high degree of precision, categorize words into classes, recognize word-forms such as in new contexts, and start to get a grasp on how words are used and combined to relay meaning. How these increasingly complex sets of linguistic abilities interact with one another is, to a great degree, an open question.

The summary below sets the groundwork to look at one piece of these interactions: how word learning fits in with the other language abilities that are developing in infancy. To understand word learning in this context, it is necessary to lay out the linguistic abilities across different language areas that emerge in this time range. These abilities, in turn, allow infants to recognize and eventually recover meaning from the word-forms they hear. To do this successfully, infants must learn to identify various linguistic units in the utterances they are exposed to (e.g., words, morphemes, clause boundaries, etc.). They must also learn to correctly interpret non-phonemic variation, such as that due to prosodic, affective, and voice-quality differences. These skill sets together help the infants solve the word-form problem. While I make reference to word comprehension to some degree in this section, I leave a discussion of the theory and research in that domain for the section on the mapping problem below.

**Initial language abilities.**

Infants are born equipped with an auditory system that can already make many impressive linguistic distinctions. Some of these abilities are derived from their linguistic experiences *in utero*. For example, newborns prefer their language to other languages (Mehler, et al., 1988) if the other language is not too similar to theirs (Nazzi, Bertoncini, & Mehler, 1998). They also prefer their mother’s voice over the voice of other women (Decasper & Fifer, 1980).

Other very early language abilities seem to be present regardless of the specific content of these prenatal experiences (though they too could in principle rely on there being linguistic input of some kind). For example, newborns can discriminate many of the sounds of the worlds’ languages, whether or not their language uses a given distinction (Eimas, Siquelan, Jusczyk, & Vigorito, 1971; Werker & Tees, 1984). This ability is not as obvious as it may sound: for instance, discriminating between a /b/ and /p/, or any other pair of speech-sounds that varies in voicing alone, requires non-linear discrimination along a *continuous* dimension, namely voice onset time (VOT; when voicing of the vocal folds occurs in relation to when the closure in the mouth is released).
That infants discriminate /b/ and /p/ but not two instances of /b/ that have the same VOT difference is often taken as evidence not just of nonlinear discrimination, but of categorical perception (Eimas, et al., 1971). Newborns can also tell apart lists of function words (grammatical words, e.g. prepositions) from lists of content words (lexical words, e.g. nouns) in their native or in other languages, based presumably on the sound property tendencies of these words as a group (Shi, Werker, & Morgan, 1999).

These very early abilities within the speech-sound domain demonstrate that at the very beginning of language acquisition, infants learn about a specific level of linguistic representation, namely the sound level, quite apart from learning within other linguistic domains such as semantics and syntax, which have yet to develop. Moreover, the mix of early abilities shows sensitivity to a swath of acoustic dimensions: speaker identity, phonemic discrimination, and sound properties (e.g. stress and prosody) of groups of words. These abilities are built upon by later language learning: not only do infants need to figure out which kinds of acoustic differences there are in (their) language, they will also need to discover which of these are relevant for meaning (e.g. phoneme differences as in ‘bear’ and ‘pear’), and which are not (e.g. voice quality differences between individuals).

It is also worth bearing in mind that early speech experience does not occur in a vacuum, independently of other kinds of cognitive stimulation. Infants hear language while also taking in a multitude of other stimuli: they see, touch, and taste elements of the world around them, and interact with their caregivers. That is, given the redundancy across different modalities and modes of communication, it is very likely that when infants learn about speech sounds, they do so by noticing how speech sounds interact with other aspects of perception and cognition as well.

**Language-specific phoneme discrimination.**

From this broad set of early discrimination abilities and preferences, infants begin to winnow down the input into language-relevant categories. This process may be helpful for isolating word-forms, if infants’ language-general discrimination abilities are not sufficient for this task. While debate continues about the mechanisms through which language-specific phoneme narrowing occurs, researchers generally agree on an account along the lines of Kuhl (2000): infants learn about their native language speech sounds by finding patterns in their linguistic input and, building from the statistical properties within these patterns, find language-specific speech-sound units. This process leads to a ‘perceptual magnet’ effect whereby infants’ experiences with speech sounds warp their perception of them (Kuhl, 2000).

This account is supported by data showing that between six and twelve months, infants’ speech-sound discrimination narrows to just those categories that are relevant for their language, both for vowels (Bosch & Sebastian-Galles, 2003; Polka & Werker, 1994) and for consonants (Werker & Tees, 1984). For example, while English learning 6-month-olds can discriminate a contrast between dental and retroflex /t/ found in Hindi, by their first birthday they can no longer do so, while Hindi hearing infants maintain their ability to discriminate these sounds; vowels seem to be ahead of consonants in this narrowing.
One proposed mechanism for how infants determine their language’s specific phonemes is distributional learning. That is, languages split up the acoustic space differently, such that in the continuum from ‘t’ to ‘d’, which varies in Voice Onset Time (VOT), a language may have a single category (as in Maori), two categories (as in English), or even three categories (as in Armenian). With any number of categories, the tokens infants hear vary to some degree, but how they differ in their distribution of VOT is systematic, depending on the underlying number of categories. Maye, Werker and Gerken (2002) showed that infants are sensitive to this sort of distribution. They familiarized 6 and 8 month olds with either two minutes of ‘t’ and ‘d’ sounds with bimodally distributed VOT, or unimodally distributed sounds; infants were able to discriminate the endpoints of the continuum only if they’d received the bimodal familiarization. Similar findings come from Kuhl and colleagues, who show that exposure to the prototypical vowels within a language exerts a perceptual pull, leading infants to categorize vowels in a language-specific way by 6 months (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992).

However, there is debate in the literature about whether the distributional cues in the input are sufficiently strong to account for infants’ acquisition of, for instance, a given language’s vowel system. The argument is that the vowel system of English (and many other languages) is just too grossly overlapping for infants to be able to pull apart its categories from their distributions alone. Recent suggestions include the possibility that infants use familiar word-forms to help them learn these categories (Feldman, Myers, White, Griffiths, & Morgan, 2013; Swingley, 2009).

Thus, between six and twelve months, infants’ phonology is still settling, at least in the case of consonants. From twelve months onward, infants have more or less solidified their native-language phoneme inventory. That said, certain speech sound contrasts do appear to show a protracted developmental timeline, most likely due to the difficulty of the perceptual discrimination under question. For example, word final voicing contrasts are still undetectable by Dutch infants at 16 months (Zamuner, 2006). This case is particularly notable, because in Dutch, word-final devoicing for stops and fricatives is a phonological rule, and occurs very systematically and predictably.

Why should language-specific phonemic inventory narrowing be necessary? Wouldn’t it be just as useful, if not more so, to retain the ability to detect all potential phoneme differences in any language? While perhaps a disappointing loss of skill, especially to the second language adult learner, phoneme inventory narrowing plays an important role in word-learning. Namely, knowing that words have different sounds makes it far more likely that they have different meanings; a better grasp on which sound-differences are phonemic in an infants’ native language may help with this learning process. For example, ‘ray’ and ‘lay’ in English have two distinct onset consonants, and different meanings, while ‘rei’ in Japanese, whose onset consonant can sound like ‘r’ or ‘l’, maps onto one word, ‘ghost’. If Japanese-learning children maintained their ability to differentiate ‘l’ and ‘r’, which at 6-8 months is

---

1 Here I put aside the case of homophones.
indistinguishable from English-learning infants’ ability (Kuhl, et al., 2006), they might find it harder to learn the meaning of words like ‘rei’. Thus, the suggestion in this line of reasoning is that retaining discrimination that is irrelevant for the native language may interfere with word learning. Indeed, Kuhl and colleagues (2008) find that while better native-language phoneme discrimination at 6 months is linked to larger subsequent vocabularies, as determined by parental checklist, better non-native speech sound discrimination actually showed the opposite pattern: infants with a better ability to tell apart speech sounds that are not relevant for their native language were reported to know fewer words in their native language 7-10 months later. This work suggests that language-specific phoneme inventory narrowing may be a crucial prerequisite for word learning. At the same time, it may turn out to be the case that the generic set of phonemes infants have early on are sufficient for word-learning to get off the ground.

**Speech invariance.**

While the kinds of discrimination described above become crucial for relaying word reference, other kinds of speech differences are never used to change a word’s identity, such as whether a word is spoken happily or sadly. Infants must learn this sort of generalization as well. That is, they must learn that when understanding words, the acoustic differences that map onto indexical properties should be ignored.

The question of how competent language users know that two utterances are the same, across different acoustic realizations, is known as the speech invariance problem (Perkell & Klatt, 1986). There are two broad areas within this problem: individual differences, e.g. from different lengths of vocal tract, different voice quality, emotions, etc. that change the acoustics of an utterance; and sound differences conditioned on phonetic context, e.g. VOT, word-position of a phoneme, etc. (i.e., allophonic differences). Language learners must come to appreciate both of these kinds of speech invariance, and the groundwork for this learning is set in infancy.

Within the speaker-difference domain, one set of studies has focused on infants’ discrimination of affect. Singh, Morgan, and White (2004) exposed 7.5 and 10.5 month olds to isolated words spoken in happy or neutral affect. Then, the words were embedded in passages that either maintained the affect from the familiarization, or switched affect. Younger infants only preferred the familiarized words if the affect was constant across the training and test phases, while older infants preferred the trained words regardless of affect. The same developmental timeline is found in infants’ preference of trained words that vary in the gender of the speaker (Houston & Jusczyk, 2000). These studies show that around 10 months, but perhaps not sooner, infants develop sensitivity to a set of indexical vocal properties that do not carry meaning, but are useful for determining other

---

2 Many phonetic properties can go either way. For instance, length is phonemic in Finnish and Japanese, and tone is used phonemically in Thai and Mandarin, but neither of these properties would create minimal pairs between words of English. These cases are not critical for the current discussion, except to highlight the fact that which properties are relevant for meaning is not entirely transparent from the nature of the property.
aspects of an utterance: that is, who’s talking to you, and what emotion their words seek to convey.

Within the second area of speech invariance, recognizing speech sounds across phonetic contexts, there is far less evidence about the acquisition timeline. Initially, 2 month old infants discriminate allophonic contrasts as if they were phonemic, for instance responding differently to the ‘t’ in ‘nitraterate’ and in ‘night rate’ (Hohne & Jusczyk, 1994). This discrimination demonstrates their sensitivity to the acoustics of the realization of these sounds, but leaves open the question of when infants realize that both of these sounds are variants of the phoneme /t/, and that the appearance of various allophones is conditioned on the surrounding phonetic context.

In a recent investigation into the development of allophonic knowledge, Seidl and colleagues (Seidl, Cristia, Bernard, & Onishi, 2009) looked at vowel nasality in 4 and 11 month old French- and English-learning infants. Vowel nasality is phonemic in French but not in English (cf. pas (step) vs. pain (bread) differ phonetically only in vowel nasality, and differ in meaning, whereas in English, nasal vowels occur exclusively before nasal consonants, e.g. had vs. hand). 11 month old French- and 4 month old English- learning infants were able to learn a rule based on vowel nasality (e.g. oral vowels are followed by fricatives, nasal vowels are followed by obstruents), and extend it to novel consonants and vowels, while 11 month old English infants were not, suggesting that by 11 months, English-learning infants have learned that vowel nasality is conditioned on phonological context. This case is in a way analogous to the narrowing phoneme inventory described above: while infants are initially flexible to interpret nasality phonemically or non-phonemically, before their first birthday they have deduced, in the case of English, that vowel nasality is entirely predictable by phonetic context, and thus poor fodder for a phonological rule.3

Note that in both cases of speech invariance described above the challenge for the infant is not to tell apart two utterances that vary in, e.g. affect or nasality, but rather to generalize across these properties and treat these instances equivalently. This is a much harder task: given that there are acoustic differences to be noticed, generalizing over them to decide that input is ‘the same’ on a given dimension requires a high degree of abstraction over the auditory input. However, even in the cases just described, infants’ task was to discriminate between two sets of auditory stimuli: one that bore a relevant relation to the short in-lab familiarization they were exposed to, and one that did not. This may be quite a different matter than using these sorts of differences correctly when understanding words: we will return to this issue in Chapter 4.

To summarize, in the domain of speech invariance we find that by around 10-11 months infants have learned to abstract over certain speaker-specific dimensions, and have begun to appreciate phonologically conditioned variation as well, showing a more sophisticated understanding of the phonemic and non-phonemic differences of their

3 If these developments are indeed analogous, one might expect that language-specific allophonic processing too would correlate with word learning; future work is needed to test this possibility.
native language. We now turn to another important linguistic prerequisite for learning word meaning: finding words in the speech stream.

**Word segmentation and word form recognition.**

Figuring out which bits of a stream of speech are words is no small feat for infants, although this intuition comes easily to adults hearing a language they are beginning to learn or have never heard before. This difficult problem is often construed as two separate abilities: figuring out where the ‘spaces’ between words are, i.e. word segmentation; and recognizing certain bits of speech as familiar chunks, i.e. word form recognition. Note that in principle, these abilities could be decoupled, that is, one could determine where word, clause, and sentence breaks are without recognizing the words, and one could recognize a word only when it is in isolation, but not when it is in a stream of speech. However, the research on this topic suggests that this is not how language learning proceeds. Rather, segmenting the speech stream and recognizing words within it seem to go hand in hand. Moreover, there seem to be several different cues that allow infants to separate and recognize words, some of which are language-general (e.g. prosody, statistical regularity and systematicity), some of which are language-specific (e.g. stress, phonotactic constraints), and some of which use what the child already knows to help her find other words in the speech stream (see Johnson & Jusczyk, 2001; Swingley, 2009 for further discussion).

**Using known words as anchors.**

One way that infants find words in a speech stream is by using familiar word-forms to segment and learn adjoining ones. One of the most salient strings of sounds to infants is their own name. Infants recognize their name, and are not misled by alternatives that have the same number of syllables and stress pattern by 4.5 months of age (e.g., baby Joshua prefers to listen to someone saying ‘Joshua’ than someone saying ‘Agatha’) (Mandel, Jusczyk, & Pisoni, 1995). Soon thereafter, by 6 months, infants use their name to pull out nearby words (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). That is, upon hearing, “Hannah’s cup was red and shiny,” baby Hannah prefers to listen to the word ‘cup’ over another word that had been preceded by another name (e.g. “Maggie’s bike had big black wheels.”) Thus, infants recognize their own name, and use it to isolate other words in the speech stream in the first half year of life. The same holds true of the word ‘mommy,’ which at this same age, infants have not only isolated as a word, but have invested with meaning as well (Tincoff & Jusczyk, 1999).

A similar, but more complicated pattern is found with content and function words. As mentioned above, infants can discriminate lists of content and function words at birth, suggesting they are sensitive to the broad phonetic properties of these word classes (though not necessarily that they recognize or could classify specific words in each class). By six months, infants prefer to listen to lists of content words over lists of function words, even if the words are not in their language (Shi & Werker, 2001, 2003). A few months thereafter, between 8 and 13 months, infants’ use of function words becomes language-specific: they begin to use known, frequent function words to help them isolate and then recognize the novel content words that follow them (Shi & Lepage, 2008; Shi,
Werker, & Cutler, 2006). Thus, by around their first birthday, infants can use function words in their native language to pull out further words from the speech stream.

In each of the cases just described, infants use a known form, i.e. their name, their word for mother, or a set of function words, to help them find other word forms. However, this description of word segmentation is misleading in two ways. First, it is not clear that these two cases are analogous: an infant’s own name and the word ‘mommy’ are highly trained, precisely stored, stressed, and salient labels in the child’s environment, whereas function words are short, unstressed, not salient, and for many months to come, bereft of any ‘meaning.’ What these types of words have in common is that they are very frequent in the child’s linguistic input, and this property may be what allows them to be used as anchors. However, the several months’ delay between when proper nouns and function words are used as anchors may be due to the differences in these word classes just outlined.

The second way in which this picture of word segmentation may be misleading is that it seems, at first blush, circular: infants need words to find words. However, it seems plausible that early proper nouns are so salient and special in infants’ experiences that these are exactly the words that help infants break into this loop. A few months thereafter, function words too have become so frequent in the input that they form good anchors as well.

Once infants have entered this use-words-to-find-words loop, they become skilled at finding words in fluent speech rapidly. For instance, Jusczyk and Aslin (1995) found that by 7.5 months (but not at 6 months), infants preferred isolated words from short passages they’d heard, and preferred passages containing words that they’d just heard in isolation, and that they did so with phonetic precision. Indeed, it’s possible to conceive of infants’ success in this work as stemming from their growing ability to use function words as anchors: in four of the six sentences used in Jusczyk and Aslin’s passages, the key word (bike, feet, cup, or dog) was preceded by a function word.

Using known words to find new words is not infants’ only segmentation strategy: they have a set of language-general and language-specific cues in their toolbox early on. While the anchoring mechanism just discussed is in principle also a language-generic cue, given that there will be more frequent and familiar word-forms for infants in every language, this cue requires a proto-lexicon of learned forms, unlike the use of transitional probabilities and edges to find word boundaries discussed below.

Language-generic cues.

Word segmentation would be much easier if each word occurred in isolation frequently. Indeed, when word-forms do occur in isolation, infants seem better able to learn them (e.g. Brent & Siskind, 2001; Jusczyk & Aslin, 1995). Unfortunately, words do not occur in isolation very often, even in child directed speech (van de Weijer, 1998; J. Z. Woodward & Aslin, 1990), and this is especially the case for certain classes of words (closed class words, and in English, excepting commands, verbs). However, infants do quickly tune in to the brief pauses at the ends of sentences and clauses that are a robust language-general indicator of word boundaries (and more common than words in isolation), taking advantage of this prosodic cue. For example, by 6 months, infants
prefer speech with pauses at clause boundaries in their native language, but not in other languages (Hirsh-Pasek, et al., 1987). By 8 months, infants are better able to pick out a match between words heard in isolation and in passages if the words occurred passage-initially or passage-finally (Seidl & Johnson, 2006).

Patterns of syllables are also a strong, language-general cue that infants can use to segment the input, by, for instance, tracking specific patterns of frequently co-occurring syllables (e.g., the syllables ‘tee-bay’ occur less often than ‘bay-bee’ in English, though both are in the sentence ‘pretty baby,’ suggesting that the latter is a word but the former is not.) Pattern segmentation is a robust learning mechanism more generally, and applies to visual shapes and other types of auditory input (Kirkham, Slemmer, & Johnson, 2002; Saffran, 2003; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999). By 8 months, infants use this cue when listening to a two-minute series of syllables, and subsequently discriminate a series of syllables that co-occurred reliably in the familiarization over those that occurred less reliably, or not at all (Pelucchi, Hay, & Saffran, 2009; Saffran, et al., 1996).

Recent work (Jarosz & Johnson, 2013) suggests that language-general cues such as those discussed above are of varying helpfulness in segmenting words cross-linguistically. Moreover, English may be particularly well-segmented using these cues, as compared to languages like Turkish and Polish, which have very different phonologies, morphologies, systems of stress and prosody, etc. Future work on large cross-linguistic child-directed corpora is needed to evaluate the usefulness of cues currently described as highly reliable and language-generic, to determine the degree to which they are actually useful across languages. The same applies to language-specific cues, described below.

**Language-specific cues.**

Language-specific word-segmentation cues, such as phonotactic patterns, also provide a cue for infants about word boundaries. Between six and nine months, infants tune in to which sounds in their language make for good word beginnings, leading to a preference for licit word onsets over illicit ones (e.g., in English, ‘st’ vs. ‘zw’ (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). Also, certain sound sequences are more likely to span word boundaries than to occur within words, e.g. ‘sd’ is far more frequent across words, e.g. ‘this dog is nice’ than with in words, e.g. ‘I have disdain for him.’ Using these sorts of phonotactic regularities, Mattys & Jusczyk (2000) find that by 9 months, infants use the likelihood of sounds occurring between versus within words to segment words. That is, English-hearing infants prefer to listen to a nonce word (e.g. ‘tove’) that occurred bordered by sounds that are very likely only between words in English (/vt/ as in ‘a brave tove trusts’), over a nonce word that occurred bordered by sounds that are very likely within words (/ft/ and /vn/as in ‘a gruff tove knows’) (Mattys & Jusczyk, 2000, Appendix A).

A second language-specific cue infants use to segment words is lexical stress, which is a good indicator of where words begin and end, at least in languages like English and French where the stress is highly systematic (initial in the former, final in the latter). Moreover, once infants have appreciated their language’s stress pattern, they not only come to prefer it for words they do not know, they are also better able to pull out
words following the dominant pattern than words that diverge from it. For example, English-learning 9 month olds (but not 6 month olds) prefer to listen to words with trochaic stress (e.g. paper) over words with iambic stress (e.g. guitar), even when these words are unfamiliar to them (Jusczyk, Cutler, & Redanz, 1993). Around this same time, at 7.5 months, English-learning infants hearing passages containing words with either trochaic or iambic stress only prefer the key words in isolation when they are trochaic (matching their language), though by 10.5 months they are able to recognize the iambic words as well. This is an important development, because although it is not the dominant sound pattern in English, plenty of words in English are iambic, e.g. ‘hooray’ (Jusczyk, Houston, & Newsome, 1999).

Thus, there are many kinds of cues, of which I’ve just discussed five, that allow infants to break up the speech stream into words. Given that these cues are a mix of those that would be informative in any language, and those that are only informative in the native language, it has been suggested that the generic cues pave the way for the language specific cues (Johnson & Jusczyk, 2001; Swingley, 2009; Thiessen & Saffran, 2003). However, the ages at which various cues have been demonstrated suggest that infants are likely using multiple cues in parallel, though perhaps the weighing of generic versus specific cues is part of what develops in the second half of the first year.

Each of the cues discussed above is not a hard and fast rule. For instance, using the word ‘mommy’ or a clause boundary as an anchor is helpful for knowing where a word begins, but less so for knowing where it ends. Stress and transitional probabilities are more useful for parsing out bisyllabic words than monosyllabic words (or the polysyllabic words found in agglutinative languages, like Hungarian, see Gervain & Mehler, 2010). Phonotactics too often reflect tendencies rather than rules, especially when allowing for proper nouns to be part of the growing lexicon. Given these facts, it is no surprise that the infant would need to recruit many of these cues in parallel.

Another possibility is that once infants understand a few words, they can use meaning to boost the accuracy of these other segmentation cues: if the word suggested by these other cues has semantics, or even the potential semantics given the environmental context in which it is heard, this may in turn boost the likelihood that it is a word. That is, an infant hearing, “Look at Mommy’s apple! Isn’t it nice?” could use stress, transitional probabilities, word boundaries, and phonotactics to hypothesize that ‘apple’ is a word, and if there is indeed an object in her view that she knows to be or could consider a viable referent for apple, she can be all the more sure that this chunk is a standalone word. This type of strategy does not help the infant with ‘at’ or any of the words in the second sentence, but it could still be a step towards better segmentation.

**Word-form precision.**

One may wonder how precise the phonetic detail is in the words that infants are pulling out of the speech stream. In the Jusczyk and Aslin study discussed above (1995), infants only segmented out the key word from the passages if in both the passages and isolated forms it was pronounced in the same way—‘gike’ was not accepted by infants as ‘close enough’ to lead to recognition of the word-form ‘bike.’ Of course, these passages were not likely infants’ first exposures to the tested words (bike, cup, dog, and feet), but
the study did indicate that changing the initial phoneme was enough to disrupt infants’ recognition of these word-forms; a fuzzier, more forgiving representation may have resulted in a false positive.

Subsequent research has found that indeed, infants do store word-forms faithfully, and are sensitive to mispronunciations. For instance, at 11 months, Dutch-learning infants prefer correctly pronounced words over words that are mispronounced in their initial sound (Swingley, 2005). A few months later, at 12-14 months, infants show better performance when linking correctly pronounced words with images depicting them, than when the words are slightly mispronounced, e.g. ‘gall’ for ‘ball’ (Mani & Plunkett, 2010; Swingley & Aslin, 2002). These studies, among others, provide compelling evidence that infants have precise word forms in their mental lexicon, early in development in terms of their word-form recognition, and later on in terms of their ability to link word-forms and referents. It remains an open question whether word comprehension at its earliest stages too shares these precise phonetic representations (see Chapter 4).

The research on how precisely word forms are stored in infants’ memories contains an implicit assumption that words in infants’ lexicons go through a phase in which they are stored as word-forms, devoid of meaning. This has been demonstrated in the lab with toddlers (Swingley, 2007), but it is not yet clear how this process unfolds with infants’ early words.

**Beginnings of grammatical knowledge.**

While there is little reason to think that infants’ early words belong to adult syntactic categories like ‘preposition’ or ‘noun,’ there are some indications that in the first year infants are beginning to gather the necessary knowledge that will turn into these syntactic categories. Some of the results concerning function and content words were discussed above (e.g. Shi, et al., 2006). Other work by Gervain and colleagues has investigated whether infants realize their language is head-final or head-initial (grossly characterized, whether function words precede or follow content words; languages vary along this dimension, e.g. Italian is head-initial while Japanese is head-final). In this work, 8 month olds were exposed to a series of nonce function and content words, such that it was not clear whether the function words preceded or followed the content words; Japanese learning infants preferred a content-before-function pattern of test syllables, i.e. a head-final pattern, over the opposite. Italian-learning infants preferred the function-before-content, i.e. head-initial pattern, suggesting that infants at this age are sensitive to where functional elements tend to occur in their language.

Patterns over syllable series are also a robust, language-general cue that infants can use to learn the building blocks for grammatical information, e.g., that simple English sentences typically follow a noun-verb-noun pattern. At 7 months, infants are able to learn an abstract pattern (e.g. A-B-B) instantiated over syllables (e.g. la-we-we, ro-fa-fa), and differentiate it from a new pattern at test (e.g. ki-so-ki vs. ki-so-so). This type of representation in turn, may become useful for further syntactic and morphological learning and categorization once abstract word-class categories emerge.

Finally, by 15 months, infants already track aspects of the argument structure in their native language. Jin and Fisher find, in recent work (Jin & Fisher, 2013), that 15
month olds learning English expect a transitive scene with two actors when they hear ‘he’s gorping the boy,’ but an intransitive scene (i.e. with one actor) when they hear, e.g. ‘He’s gorping.’ In the few months thereafter, infants’ understanding of word-order constraints and transitivity continues to expand (Gleitman, 1990; Golinkoff, et al., 1987; Yuan & Fisher, 2009).

**Language production.**

In the first year of life, infants’ language production mirrors some of the findings we’ve seen in language comprehension. Infants transition from producing non-speech sounds (0-3 months), to producing vowel-like sounds around three months, to beginning canonical babbling around seven months (e.g. ‘ba,’ ‘da’). These early developments have to do with physiological constraints (Goldstein & Schwade, 2010), rather than particular language exposure, but they soon give way to language-specific speech sound production around 10 months (de Boysson-Bardies, 1993; de Boysson-Bardies, Halle, Sagart, & Durand, 1989), with the first words usually uttered around infants’ first birthday (Kuhl, 2004, fig. 1). Thus, research on the development of speech production in the first 10 months has shown how infants gain control over their speech production system, articulators, etc. and gradually shift from language-general to language-specific sounds, and soon thereafter to word-forms with a consistent form and meaning.

Situating language production in a social context, Goldstein and colleagues have conducted a series of studies looking at infants’ vocalizations during infant-mother interactions. They find that infants’ vocalizations change as a function of their mothers’ responsiveness (Goldstein, King, & West, 2003). In this work, mothers played with 8-10 month old infants and gave nonverbal support (smiling, touching them, moving towards them) that was either time-locked to their child’s vocalization, or time-locked to another child’s, and thus out-of-sync with their own child’s vocalizations. In this context, only the infants whose mothers responded contingently showed advanced vocal productions (e.g. faster consonant-vowel transitions), which carried over into a later play-period (Goldstein, et al., 2003). In a similar study, again, only infants whose parents had responded to them contingently increased the type of vocal structure they exhibited, but here they did so as a function of whether their mothers repeated their babble as a babble, or repeated it back as if it were a word, suggesting that not only are infants’ productions influenced by their mother’s feedback, but by aspects of the structure of her vocalization as well (Goldstein & Schwade, 2008).

While infants’ responses in relation to their parental feedback are interesting in and of themselves, and confirm that by 8-10 months infants are active social partners, this line of research has also tied infants’ vocalizations to later word knowledge. In another study (described in Goldstein & Schwade, 2010), mothers played with their 9-month-olds, and mothers’ responses to children’s object-directed vocalizations were measured. Six months later, infants’ vocabulary was measured using a vocabulary checklist, the Macarthur Communicative Development Inventory (CDI, Fenson, et al., 1994) known to show long-term predictive validity for infant language development (Fenson, et al., 2000). In the playing phase of the study, parent behavior after their child vocalized towards an object fell into two categories: ‘proximal object labeling,’ in which parents
said the name of the object their child was interacting with, and ‘phonological resemblance,’ in which parents said a word that sounded similar to the vocalization the child had made. The proportion of proximal object behavior correlated with later productive vocabulary, while the proportion of phonological resemblance behavior correlated with a smaller later comprehension vocabulary (described in Goldstein & Schwade, 2010). These results are interpreted by the authors as showing that the type of vocalization parents respond with when their child is attending to and vocalizing toward an object has repercussions on their child’s subsequent word learning. However, this study raises some questions about the mechanisms underlying infants’ learning from their mother’s responses: why would increased proximal object labeling lead to an increase in production while phonological resemblance labeling lead to fewer words being understood? Future work will need to look at this question.

Another study in this line of research compared 12 month olds’ recognition of objects they vocalized more towards and those they vocalized less towards, and found that recognition was better to the former (importantly, the time spent looking at and holding the familiarized objects did not differ Goldstein, Schwade, Briesch, & Syal, 2010). In a subsequent study, 11.5 month old infants showed better learning of a novel object label when the word for it was presented contingently with a vocalization from the child towards the object (Goldstein & Schwade, 2010).

Infants do not produce many vocalizations that would count as the words they’re learning to segment and understand in the first year of life. However, their growingly sophisticated language production abilities demonstrate that their vocalizations towards objects, and how their mothers respond to these vocalizations, have a role to play in determining infants’ ability to learn words and thus in their vocabulary size at 12-15 months.

**Word-form problem summary.**

In this section we have seen that finding the words in the speech-stream is a multifaceted and difficult problem for infants to solve. They are aided by innate or early-learned abilities to discriminate speech sounds in a surprisingly astute way, and this ability is honed over the first year of life, leaving infants with the appropriate phoneme inventory for their language between 6 and 12 months. At the same time infants are learning not just what phonemes are in their language but about non-phonemic speech sound differences as well. For instance, they learn that though context alters phonemes’ acoustics, and affect, gender, and other individual differences alter how words sound, these types of differences matter in a different way than those between a ‘p’ and a ‘b.’ While these properties of the language’s speech sounds are being learned, infants are also beginning to separate the speech stream into words, using a diverse collection of cues that are both language-generic and language-specific, and some that build on other words infants have already isolated, such as their names.

As I’ve suggested along the way, knowing what even a small set of words mean could be very useful in helping confirm what infants have deduced from their analysis of the speech-stream alone. While it is now clear that understanding minimal pairs (e.g. bear, pear) is not how infants learn speech sounds, nevertheless, knowing some
words could increase infants’ confidence that words they’ve segmented, or phonemes they’ve posited, are indeed correct. Having this groundwork in place early on could help explain the speed with which language learning takes off after infants’ first birthday.

The Concept Problem

Everyone can agree on reasonable criteria for what it means for an infant to ‘find a word,’ i.e., do they recognize the word form in an appropriately broad (e.g. across speakers, times of day, contexts) and narrow (e.g. correct sounds and stress are required) set of circumstances? What it means to ‘find a concept’ is a far trickier matter. Indeed, philosophers point out that we can never be sure that the concept we have in mind is the same one that others do (Quine, 1960), and they disagree about whether the meanings of concepts like ‘dog’ or ‘gold’ are to be found in the world or as individual mental entities (e.g. Fodor, 1998; Fodor, Garrett, Walker, & Parkes, 1980; Putnam, 1975).

There have been many proposals for how concepts are structured. To the British Empiricists, concepts were combinations of innately given ‘simples’ that in turn were directly available through general perceptual mechanisms (Hume, 1739/1978; Locke, 1690/1975). Proposals over the last forty years have included the following possibilities: that having a concept means knowing its extension (all possible members of the class), or intension (a rule that picks out only and all the right members), knowing its necessary and sufficient features, having a probabilistic knowledge of its features, having theories about its essence, and appreciating it as an unbroken whole that is innate or immediately appreciable by perception (Armstrong, Gleitman, & Gleitman, 1983; Carnap, 1947; Fodor, 1998; Fodor, et al., 1980; Fodor & Lepore, 1996; Keil, 1994; Rips, 1989; Rosch, 1978; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; E. E. Smith & Medin, 1981, inter alia). See discussion in L. B. Smith (2005) of some of these views). None of these proposals has withstood scrutiny, on both theoretical and empirical grounds, nor do many of them suggest how the infant learner could come to acquire concepts.

Rather than reviewing and analyzing these various ideas of what a concept is (though see Bergelson, 2011), I want to focus on the commonality that these theories try to account for: we are able to communicate, by and large without problem, in reference to all sorts of concepts, from ‘penguin’ to ‘iridescent’ to ‘noshing.’ This ease in communication suggests that whether or not we can spell out what a concept means, when we use a content word, we use it to refer to concepts that are close enough to others’ concepts to lead to successful communication.

Turning to the original question, how does the infant come to appreciate the concepts that are to be picked out by words? One way to get at this is to ask what infants know about the world around them. Research in the past 30 years has demonstrated that infants have an impressive grasp of the physical world, and of how individuals in that world tend to behave. Seminal work by Leslie, Spelke, Baillargeon, Carey, Woodward, and others has led to incredible growth in our understanding of what knowledge infants are born with, and how this knowledge develops over the first few years of life. The relevance of this work to the current dissertation is two-fold: on a more abstract level, some of the debates about the mechanisms and origins underlying infants’ reasoning apply to word-learning as well; and on a more concrete level, what infants know or
reason about objects, events, and agents is a crucial part of how they link words to these aspects of their experience. Thus, a summary of the relevant literature helps ground our understanding of the word-learning task, and helps set the stage for the questions addressed in the empirical work in the chapters to come.

**Domains of infant reasoning.**

The research of Spelke, Baillargeon, and others in the past few decades has brought to light infants’ impressive skills within many domains, upon which other skills and beliefs are built. These are often dubbed ‘core systems,’ though here we do not delve into the particular implications of this terminology. These domains of research include a representational system for objects, agents and actions, number, space, and social partners (Spelke, Breinlinger, Macomber, & Jacobson, 1992; Spelke & Kinzler, 2007, inter alia). Each system has elements that can be individuated and used for forming inferences. In a sense, aspects of these abilities are analogous to the initial language abilities described in the preceding section: they do not seem to require very much exposure or learning before they show themselves, which suggests that these abilities make use of mechanisms, representations, and biases that are part of the cognitive endowment of the child. However, these (let’s say innate) abilities are not enough for acquiring adult-like facility in a domain, which requires experience with the environment and agents within this environment.

Baillargeon and colleagues propose a similar set of divisions in infants’ reasoning abilities into physical, psychological, and biological domains (see Baillargeon, Li, Gertner, & Wu, 2011; Baillargeon, et al., in press for reviews). For ease of exposition I use these sets of divisions loosely, splitting the research into knowledge and reasoning about objects, and knowledge and reasoning about agents, and adding in relevant work that is not within Spelke or Baillargeon’s framework, strictly speaking, but is potentially relevant for word-learning.

**Physical reasoning.**

Infants’ ability to represent and reason about objects as such, and about objects as individual members of a class, are both important abilities, maybe even prerequisites, for word-learning of object labels. That is, if infants lacked the ability to tell objects apart, or thought that every time an object reappeared in their view it was a different object, the referent for word-forms they identified would be a moving target. Thus, infants’ precocious understanding of how objects behave in the world, and of how different object kinds are categorized, perceptually and later functionally, is important for their subsequent mapping of words to referents.

**Object representation and reasoning.**

In the domain of object representation, research has found that before infants reach 6 months of age, they demonstrate object permanence: infants expect objects to continue to exist even when they are no longer visible, in contrast to early theories (e.g. Piaget, 1954) on this issue (e.g. Baillargeon, 1986, 1987; Baillargeon & Devos, 1991; Spelke, et al., 1992). Infants also exhibit expectations in line with spatiotemporal
principles: they expect objects to stay in one piece (cohesion), to move together smoothly in space (continuity), and to only act upon each other if touching (contact) (Spelke, 1990). By 3 months, for instance, infants’ looking behavior at visual scenes is consistent with an expectation that an object that moves behind an occluder comes out the other side, and is visible if there is a window in the occluder (Aguiar & Baillargeon, 1999).

In contrast to this research with object representations, which infants exhibit understanding of within the first few months of life, other categories of inanimate entities, such as foods and non-solid substances, even by eight or nine months, infants fail to make appropriate inferences and generalizations (Shutts, Condry, Santos, & Spelke, 2009). This suggests that certain types of entities may be privileged in infants’ physical reasoning. Whether such entities, e.g. bounded objects, are also privileged in infants’ word-learning is an open question, addressed in Chapter 2.

**SORTAL CONCEPTS AND OBJECT INDIVIDUATION.**

In order to learn the word that maps onto a concept, infants need to be able to represent and identify categories in the world around them. To this end, it is helpful to examine what infants know about sortals. Sortal concepts are those that identify kinds, or sorts of entities (Macnamara, 1982). In English, and other languages with a count-mass distinction, sortals map onto count nouns, e.g. ‘person,’ ‘object,’ ‘apple,’ etc. Recent work by Xu, Carey, and colleagues (e.g. Xu, 2007) suggests that infants quickly learn about sortals, which has repercussions for their ability to individuate and identify objects (In this literature, object individuation refers to the process by which we determine how many objects there are in a given event. Object identification refers to the process of figuring out an object’s properties, and attaching these to its representation).

Some of the research on sortals builds on infants’ knowledge of spatiotemporal properties. As research by Spelke and colleagues has shown (Spelke, 1994; Spelke, Kestenbaum, Simons, & Wein, 1995), by 4 months, infants shown two objects emerging from behind two occluders, one at a time, expect there to be two objects once the occluders are lowered, even if the two objects are identical, suggesting that they have represented some aspects of these objects’ spatiotemporal properties. 10 month olds too show this pattern in the two-occluder situation (Xu, 1997; Xu & Carey, 1996).

In contrast, when shown two objects (e.g. a duck and a ball) emerging one at a time from behind a single screen, 10-month olds do not expect two objects once the screen is lowered (Xu & Carey, 1996). The authors analyze this as infants knowing the sortal ‘object,’ a general category that doesn’t encode what kind of object has been seen (Xu, 2007). However, in other tasks, e.g. manual search, 10-month-olds also show an ability to represent both objects (Xu & Baker, 2005), raising a potential problem for the proposed explanation in the original study. By 12 months, infants succeed in the two-

---

4 Moreover, it’s not a lack of ability to tell the categories apart that led to 10 month olds’ performance: by 3-4 months infants can tell apart basic object kinds (Arterberry & Bornstein, 2001; Quinn, Eimas, & Rosenkrantz, 1993).
object one-screen task that 10 month olds failed at (Xu & Carey, 1996), though they too show disrupted object individuation if the objects only vary in color, size, or pattern, as opposed to kind. Interestingly, if the two objects vary only in shape, they are encoded as separate objects by 12 month olds only if the two shapes map onto two categories. That is, two differently shaped cups are not individuated, while a cup and a ball are (Xu, Carey, & Quint, 2004).

Taken together, these data suggest that infants have an appreciation of objecthood as such, which can be separated from the specific perceptual features of objects, and that in certain situations, they reason over the sortal ‘object’ rather than using the various types of object-property information they are sensitive to.

Early sortal knowledge of humans (or human-like entities) shows a different timeline: by 10 months, but not earlier, infants shown a doll emerging on one side of a single occluder and a toy on the other expect two objects once the occluder is lowered, but when shown two differently gendered doll-heads (or two dog-heads) coming out from behind the occluder, expect only one object (Bonatti, Frot, Zangl, & Mehler, 2002), suggesting that they may have the sortal ‘person’ at 10 months as well as the sortal ‘object’ (Xu, 2007). Here too, infants collapse over otherwise distinguishable features when reasoning about how many entities are behind an occluder.

Thus, the research just discussed suggests that infants have a strong ability both to focus in on the specific details of an entity (i.e., object, person, animal, etc.), such as its pattern, or to ‘zoom out’ and appreciate an object as a member of an abstract kind. This is analogous to the discussion of word-form recognition above: while infants can appreciate that a word is said in a different affect, or by a different person, they can also generalize over these perceptual differences and hear the common word. There is also perhaps a closer tie between infants’ individuation abilities and their word-learning: A link has been suggested between performance in the two-or-one task and what words infants know or hear during the task. We will return to these data in the mapping section below.

Social and psychological reasoning.

While reasoning about objects and other physical entities is relevant for learning words for them, reasoning about social and psychological processes is also helpful for word learning, though perhaps less transparently. Word learning requires learning the signifier used to symbolize a given referent in one’s linguistic community (Tomasello, 2001). Aside from rare learning instances in which infants are exposed to disembodied voices, or in recent times, iPads and iPhones, infants’ learning occurs in a social milieu. Thus, part of membership in a social, linguistic community involves understanding the intentions, goals, preferences, and actions of other individuals, and a capacity to share attention with others. In turn, aspects of these abilities may prove necessary for word learning, as examined below.

Reasoning and representation agents and actions.

Within the domain of agent and action representation, the expectations infants have for inanimate objects no longer apply. Instead, there is a plethora of evidence suggesting that infants are born equipped with a causal framework in which to interpret
the actions of agents, and attribute mental states to them (Baillargeon, et al., in press). These abilities to attribute beliefs and interpret actions may form the basis of understanding reference for word learning later in development. In short, infants expect agents to behave rationally: they expect people to perform actions that are consistent with their goals, preferences, and knowledge (consistency), and to spend minimal effort to achieve their goals (efficiency) (e.g. Csibra, Gergely, Biro, Koos, & Brockbank, 1999; A. L. Woodward, 1998).

For example, a series of studies by Woodward (A. L. Woodward, 1998, 1999) demonstrated that by five months, infants attribute preferences for objects to agents. In one study, A. L. Woodward (1999) showed infants a scene with two objects in which a hand reached in and intentionally grasped one object, or unintentionally touched it with the back of her hand. Then, in the test phase, the objects switch locations, and the hand either grasps the old object, now in the new location, or the new object, in the old location. Infants were surprised at (i.e. looked longer at) the latter event, but only in the case that the hand was acting intentionally. This was interpreted as infants understanding, as adults do, that agents tend to show preferences for specific objects (rather than their locations), and that these preferences are manifest through intentional (but not accidental) touching. Using similar methods, Luo and Johnson (2009) showed that 6 month olds expect an agent to show a preference for one object over another only if in the initial familiarization the agent has seen both objects. In each of these cases, infants reason that people should behave consistently with their knowledge and preferences over time. Similar work (Brandone & Wellman, 2009; Csibra, Biro, Koos, & Gergely, 2003) shows that infants at 6-9 months expect agents to act efficiently, by avoiding detours when possible.

In the social domain too, infants show development in their abilities: while infants are able to attribute preferences for objects to agents as early as 3 months of age (Luo, 2011), it is not until 9 months that they attribute preferences for actions (Song & Baillargeon, 2007). Later still, they attribute property-based preferences to others (Luo & Beck, 2010). In a similar vein, young infants use grasping as a sign of preference, but not pointing or gaze; by 12 months, these too also elicit an attribution of preferences to an agent (Brooks & Meltzoff, 2002; A. L. Woodward, 2003).

It is possible that the delayed timeline for attributing preferences for actions as opposed to objects, and for understanding pointing and gaze as opposed to grasping is relevant for word-learning as well. One might imagine that acquiring the ability to attribute preferences or desires allows infants to reason about what an agent is referring to. This may help infants learn labels for objects, earlier in development, and labels for actions, later on. These abilities may be differentially used for early and late acquired words, or prove necessary for word-learning in the first place. Indeed, the assumptions that infants make about how preferences or labels are shared across community members suggest a link between the attributions discussed in this section and language learning (Henderson & Woodward, 2012). We return to this in the mapping section below.

**Reasoning and representing social partners.**
A more recently proposed system, for reasoning about social partners, bears similar findings to the other systems described above (Spelke & Kinzler, 2007), and here too one can imagine their usefulness for word-learning: social partners make for good sources of linguistic input. In the social partner reasoning domain, research shows that by three months of age, infants show a visual preference for members of their own race, but only if they are brought up in racially homogenous circumstances (Bar-Haim, Ziv, Lamy, & Hodes, 2006). By six months infants prefer someone speaking their own language over another language (Kinzler, Dupoux, & Spelke, 2007), and by their first birthday, they go a step further, preferring to eat a food eaten by someone who speaks their language over another language (Shutts, Kinzler, McKee, & Spelke, 2009).

This set of developments shows a pattern similar to those we found in language: just as infants have early-appearing speech sound discrimination that, through exposure, is winnowed down to the kind of sounds relevant for their language, so too do infants seem to come equipped with an ability to make visually-determined race-based judgments, which await their exposure to a specific race’s visual makeup. This ability is manifest early in the visual preferences infants exhibit, and is built upon subsequently in terms of their language and food preferences, which are less transparently read off of the environment than race.

**Intersubjectivity, gaze-following, and joint attention.**

Infants are social beings, and as just discussed, from an early age are able to reason about agents, actions, and social partners. However, the preceding sections glossed over some important aspects of social knowledge: intersubjectivity, gaze-following, and joint attention. Intersubjectivity is the knowledge that other people are like us; imitation is one crucial aspect of intersubjectivity, and is present at birth (Meltzoff, 2007). Gaze-following and joint attention abilities grow over the first two years of life, as infants accumulate growing experience with how their bodies, objects, and other people in the world interact. The other social skills discussed above make use of these abilities, which are important in their own right, and moreover, have implications for word-learning as well. Namely, an ability to imitate others, and co-attend to stimuli in the world may lay the groundwork for word learning.

In what Meltzoff and Brooks dub “the myth of the asocial infant” (Meltzoff & Brooks, 2007), early psychologists such as Skinner, Piaget, and Freud did not see infants as social beings, but rather as unformed creatures with reflexes but not a rich mental life. In the years since, infants were found from birth to have a representation of others, and an investment in sharing their experiences (Meltzoff & Brooks, 2007, p. 151). Indeed, newborns are able to imitate facial acts like tongue protrusion (Meltzoff & Moore, 1983, 1989). This suggests a degree of body awareness that is precocious to say the least—to imitate these sorts of movements, infants have to understand what they see and translate it onto their own bodies, with a high degree of body-part specificity, and a close pairing of action and perception that allows them to understand that others inhabit bodies like our own.

Unlike the ability to mimic facial actions, which is innate, gaze-following is a learned behavior that infants use to help them understand the intentions of others. Infants
show an early focus on the eyes, looking at them preferentially by two months (Maurer, 1985), and discriminating whether gaze is straight ahead or diverted soon thereafter (Vecera & Johnson, 1995).

However, what is really of interest here is when and how infants begin to use others’ gaze to direct their attention to the external world, since such a skill may be recruited for word-learning. That is, if joint attention is a necessary component for word learning, we might expect to find word comprehension development around the time that various aspects of joint attention and gaze-following mature.

Around 6 months, infants begin to follow another’s direction of gaze (Morales, Mundy, & Rojas, 1998). However, it is not clear if this is true ‘joint attention’ at this stage, or if the infant is simply following the adult’s head movements.

In a clever test of this phenomenon, Brooks and Meltzoff (2005) showed that infants look in the direction an experimenter is turning from 10 months onward, but only when her eyes are open. Just one month earlier (and presumably more), infants follow the experimenter’s head-turn to look at an object even when her eyes are closed. Thus, the ability to follow gaze and share attention seems to show a concrete development between 9 and 10 months of age.

Even with a more rigorous definition of joint attention, it remains possible that the attention between the infant and the object, and her caregiver and the object, is parallel rather than truly shared. This ambiguity led Carpenter and colleagues (Carpenter & Call, in press) to redefine joint attention as involving two partners that “know together that they are attending to the same thing” (p.4). So when does this truly shared form of joint attention happen, and how is it is measured? Carpenter et al. find that, again, 9 months seems to be a crucial age at which infants show more refined joint attention, in which they look at their social partner and then back at a shared event/object, and begin to produce declarative gestures whose goal is not to acquire an object or approval, but rather for their social partner to see what they are showing. A few months later, around 12 months, infants begin to use declarative points in a similar way, to indicate less proximal objects and events (Carpenter, Nagell, & Tomasello, 1998; Liszkowski, Carpenter, Henning, Striano, & Tomasello, 2004). Later still, at 14 months, infants behave in a way consistent with knowing that their attention has been shared with a specific person, in relation to a specific object (Liebal, Behne, Carpenter, & Tomasello, 2009; Liebal, Carpenter, & Tomasello, 2010).

When do infants begin to appreciate the social role of gaze in others? Beier and Spelke (2012) find the same developmental timeline as Brooks and Meltzoff (2005): 9 month olds neither discriminate between two other people in mutual or averted gaze, nor do they expect people to look at their social partners in conversation, while 10 month olds’ behavior suggests they discriminate the direction of two people’s gaze in relation to each other, and expect shared gaze between social partners engaged in conversation. The authors interpret their findings as indicative of a sharp shift in infants’ gaze understanding, suggesting that only at 10 months do they begin to appreciate that gaze is social and goal-directed.

While the mother-baby-object triangle of reference allows the infant to deduce what non-proximal object the mother’s attention is focused on without reference to
language (Meltzoff & Brooks, 2007, p. 154), researchers have suggested that gaze-following (perhaps in its less refined, not truly ‘joint’ form) is a pre-requisite for word learning. Indeed, Morales et al. find that gaze-following ability at 6 months predicts expressive vocabulary at 24 and 30 months (2000). In the same vein, Brooks and Meltzoff (2005) find that infants’ gaze-following performance at 10-11 months predicts their productive vocabulary and sentence complexity at age 2.

However, these connections between gaze-following and productive vocabulary are coarse-grained and indirect. That is, the suggestion in these studies is that appreciating gaze helps infants learn words. What is missing is a connection between early gaze-following abilities and language comprehension. Is gaze-following truly a prerequisite for word-learning, or might this depend on the kind of word to be learned? Is there a word-learning change around 9 months that would map onto the series of findings just discussed, which show a shift in gaze-following ability, using several different measures, around that time? Is the distinction drawn by Carpenter and colleagues relevant here, and if so, is there a relevant change around 12-14 months in word-learning ability that reflects this ‘truer’ joint attention? Chapter 3 begins to address these questions.

Summary of the concept problem.

The discussion presented above is intended to suggest that infants have a rich mental life, and many of the kinds of abilities that may be necessary for learning words are present early on, and built upon over the first year of life and beyond. Infants understand many aspects of how objects behave in the physical world before six months of age, and between 6 and 12 months learn to appreciate many of the elements that differentiate categories, properties, and kinds. While their reasoning about objects is particularly precocious, there are areas in which infants show immaturity, i.e. reasoning about foods and substances (Shutts, Condry, et al., 2009), and how objects and people should be differentiated (Bonatti, et al., 2002; Xu, 2007). In the domain of agents and their actions, infants show growing understanding of preference and intent over the first year (A. L. Woodward, 1999), with a seemingly sharply delineated ability to use other’s gaze coming online between 9 and 10 months (Beier & Spelke, 2012; Brooks & Meltzoff, 2005; Carpenter, Nagell, & Tomasello, 1998).

It is clear from previous research that across a range of conceptual domains, infants have an understanding of how entities and agents behave. The breadth and depth of this knowledge suggests a set of biases in infants’ cognitive system that allow them to reason about the physical, biological, psychological, and social domains quite early, and later refine this reasoning as further experiences are accumulated. The abilities described above leave no doubt that infants can individuate entities and events in the world, which is likely to be a crucial prerequisite in learning the words that signify these entities and events within a given language. How the specific links are formed between categorizing objects and properties, individuating and representing sortals, and understanding preferences and intentions on the one hand, and word-learning on the other, remains an open question. It is not yet clear which of these first sets of abilities is helpful, or even critical for successful word-world mappings by infants.
The Mapping Problem

Having seen above that infants are isolating words and individuating concepts in the first year of life, we can now turn to how they link them together. This is especially tricky considering the mapping is not one-to-one: the penguin you see at the zoo could be called ‘penguin,’ ‘animal,’ ‘bird,’ ‘creature,’ or ‘Romeo,’ and the category to which these terms extend varies widely; on the reverse end, the new dance move invented in your room at 2 AM has no word (yet!) to label it, though some superordinate terms (e.g. ‘grooving’) may apply. And yet, through a combination of their linguistic and conceptual abilities, infants soon sort this out, as we’ll see in the coming chapters, not just for nouns, but also for various kinds of words.

There is general agreement that the words infants are likeliest to comprehend first, and in the first year, are the most concrete words, which in the adult lexicon are mostly nouns. After a brief review of word comprehension research I turn to the work on novel word learning, including a discussion of the various theoretical explanations about why nouns may be easier to learn. I then turn to a brief discussion of some of the mechanisms underlying word learning, before summarizing the present and coming chapters.

Word learning.

Few studies have looked at infants’ understanding of words in their native language in the first year, mostly because it is generally assumed that most of this time is spent learning the phonetic and phonological properties of the native language, with word learning coming online only thereafter, around infants’ first birthday. What studies there are have typically focused on one of two paradigms: testing of real words, where ‘training’ comes from infants’ protracted daily experiences (word comprehension studies); and naturalistic, unstructured observational data (observational studies). There is a much larger literature looking at infants’ ability to learn novel words after short, controlled in-lab learning exposure (training studies).

Word comprehension studies.

Studies of this first kind are especially sparse in the literature. The youngest evidence of infants knowing the meaning of a word comes from Tincoff and Jusczyk (1999), in which they find that by six months of age, infants know the words mommy and daddy, as indexed by looking more at a video of the appropriate parent when these words are uttered (Tincoff & Jusczyk, 1999). Success in this task requires the infant to have picked out these words from her daily interactions with her parents, and to have mapped them appropriately onto each parent. However, these words are in the funny position of being, to the infant at any rate, proper names, and as such do not require the kind of generalization and abstraction that knowing a common noun requires. Recent work from the same authors (Tincoff & Jusczyk, 2012) shows that 6 month olds also understand the

---

5 There is some debate about whether this holds for the first words in the production vocabulary (cf L. Bloom, 1993; Nelson, 1973, inter alia).
words ‘hand’ and ‘feet.’ This is the only published study, other than the work presented in this dissertation, showing infants’ understanding of common nouns in an in-lab experiment at 6 months of age.

With slightly older infants, Thomas et al. find that infants show knowledge of common nouns like ‘dog’ at 13 months of age, but not at 11 months, in a 4-choice task in which their mother uttered either the name of an object she felt the child knew, one she felt the child did not know, or a nonce word (Thomas, et al., 1981). However, there were potential limitations in this study that may have obscured infants’ knowledge at 11 months: the set of small toy objects may have resulted in an unfairly limited set of stimuli that did not include words the child knew, and providing four choices may have been unduly challenging, especially for the younger infants. In another study, Bretherton et al. (1981) found that 13 month olds showed mixed performance when asked to manually select a named object, e.g. spoon, from among three competitors. While their performance was above chance, it is described elsewhere as a “near chance” (Bates, Bretherton, & Snyder, 1988, p. 116), and was far below the performance of 20 month olds. However, the task demands of manually selecting the correct answer may explain why 13 month olds here show mixed performance, while those in Thomas et al. succeeded.

Observational studies.

Using a more natural setting and shifting the burden of the task onto the parents, observational studies (e.g. Bates, 1993; Benedict, 1979; L. Bloom, 1993; Nelson, 1973) ask parents to keep a diary or fill out a checklist recording which words their infants know and when. While word production has generally been the focus of these sorts of studies, and thus they often don’t begin until around 12 months of age, Benedict (1979) particularly interested in the first 50 words understood and produced by eight infants, who began to be observed at 9 or 10 months. She found that by 13.5 months, on average, they had amassed 50 words in their comprehension vocabulary (using rigorous standards of what it means to ‘comprehend’), and as will become important in chapter 3, she found that these words came from a wide variety of word classes including greetings, action words, and social words, in addition to nouns, which represented 56% of the comprehension lexicon (Benedict, 1979, p. 193). Additionally, this study showed rapid early growth in the comprehension vocabulary, with less than 3 months, on average, between when infants understood 10 words and when they understood 50. One problem with this study, and most observational studies, is that they are unable to say whether children understood words before the study began. That is, if parents are only told how to rigorously assess infants’ word knowledge at the study’s onset, it may be hard for them to apply these recently learned standards to infants’ knowledge state in preceding months. This in turn may bias the reported onset of word learning to around the time the study begins.

A large-scale study using the Macarthur-Bates Communicative Development Inventory (a parental vocabulary checklist), with over 600 subjects, found that parents vary widely in the age at which they ascribe word meaning knowledge to their infants (Fenson, et al., 1994). At 10 months, for example, the lowest-scoring infants on a parental
vocabulary checklist are thought to understand fewer than a dozen words; the highest scoring infants, around 150 words. While such checklists allow for a broader breadth of data collection based on naturalistic interactions between infants and their caregivers, using them as evidence for when word learning begins and what kinds of words are first understood may be problematic. This is because it is unclear what standard of evidence parents use in deciding whether or not their child ‘knows’ a word. This is especially difficult in the case of word comprehension, rather than production, because there are few overt behavioral signs of word comprehension, and words are often accompanied by adults’ gestures and eye movements that can confound the record. Nevertheless, these checklists are a useful, if noisy, measure of infants’ word comprehension knowledge.

**Training studies.**

*Novel object, novel word.*

In-lab training studies are by far the most common form of querying infants’ word comprehension, or more accurately, their word learning. In these studies, infants are brought into the lab, usually for a single visit, and exposed to new words and objects, either live or through videos. After this familiarization phase, infants’ ability to link the new words with the new objects is queried, usually through a measure of eye-gaze among competitors, through a measure of whether infants notice that a taught name-object link has been altered, or by a manual selection task. The largest benefit of these studies is that infants’ exposure to the word and its referent is tightly controlled.

Research from several labs has shown that by 6-9 months, infants are able to learn a pairing between a novel word or syllable and a novel object (Gogate & Bahrick, 2001; Gogate, Bahrick, & Watson, 2000; Henderson & Woodward, 2012; Shukla, White, & Aslin, 2011; Stager & Werker, 1997), though each study suggests they recruit different tools to do so.

Gogate and colleagues suggest a critical role for invariance detection, which they define as “a general perceptual phenomenon whereby organisms attend to relatively stable patterns or regularities” (Gogate & Hollich, 2010, p. 496). They found that 6-8 month olds only learned object-syllable pairings when these were taught with synchrony between the object’s motion and its labeling (Gogate & Bahrick, 1998, 2001; Gogate, Bolzani, & Betancourt, 2006; Gogate & Hollich, 2010), and moreover, that looming and shaking the objects in particular facilitated infants’ learning (Matatyaho & Gogate, 2008).

Shukla et al. (2011) found that 6 month olds can parse and map a novel word onto an object. In this study, infants saw three objects, one of which was visually highlighted. At the same time, they heard a series of utterances made up of nonce syllables. Within these utterances, one nonce 3-syllable ‘word’ had a transitional probability of 1.0 between its syllables, and occurred at a prosodic phrase boundary. During the test phase of the study, infants picked out the referent upon hearing this nonce word, but not upon

---

6 An even earlier precursor to this ability is found in 2 month olds who can pair a syllable with an object, if presented synchronously but not asynchronously, in a simpler paradigm (Gogate, Prince, & Matatyaho, 2009).
hearing part-words, or a nonce word that was not prosodically aligned. This shows that by 6 months, infants are already able to pick out words, pick out referents, and link them together from a fairly realistic, but brief, in-lab exposure. Questions remain about how this maps onto word-learning of naturally occurring words in the infant’s native language, where the regularities are less regular, and the objects less delineated. That said, this study, more than most others, shows that young infants are able to do a version of just what might be required of them in their real word-learning experiences.

Henderson and Woodward show an interesting contrast in infants’ theorizing about words as opposed to other culturally relevant human behaviors (2012). 9 month olds observed an actress either expressing a preference for one of two novel objects, or applying a label to it, e.g. ‘modi.’ When shown another actor, infants did not expect the same preference to carry over, but did expect the same label to do so. This suggests that by 9 months, not only can infants link a novel word with a novel referent, as we saw in the preceding studies, but they also have expectations that this knowledge is general to their linguistic community, in contrast with their expectations about the specificity of preferences (Buresh & Woodward, 2007; Henderson & Woodward, 2012).

**Learning similar-sounding words.**

Similarly to the studies just described, Stager & Werker (1997) have shown that by 8 months, infants are able to learn a phonetic label for a novel object, and moreover, that they show surprise when the label is applied to a new object. For 8 month olds (and 17 month olds for that matter, Werker, Fennell, Corcoran, & Stager, 2002), it doesn’t matter if the phonetic labels are similar (‘bih’ and ‘dih’), but somewhat surprisingly, 14 month olds fail to learn two similar-sounding labels for two distinct objects, even though they can discriminate the two auditory word-forms in other tasks. This result has made a big impact on the field, and has led to many studies that attempt to explain why and how 14 month olds have trouble learning similar-sounding words, and more broadly, why and how phonological similarity affects word learning differentially across development.

For instance, subsequent work has shown (including further experiments in the original study) that 14 month olds are able to succeed in word learning with labels that are different enough (‘lif’ and ‘neem’), if a two-object selection task instead of a switched-labels task is used, if infants are pre-familiarized to the objects, if training phrases naming known words precede the novel word learning, or if the words are better familiarized (Fennell & Werker, 2003, 2004; Thiessen, 2007; Yoshida, Fennell, Swingley, & Werker, 2009).

These data and the discussion they’ve engendered have been very useful in clarifying what sorts of cues, learning methods, and testing methods push around performance, but nevertheless leave open the question of what changes between 8 and 14 months, and 14 and 17 months in the original task. The suggestion offered by Stager and Werker (1997) is that 8 month olds could be responding based on their discrimination of the speech sounds, without having truly formed a link between the word and the object they were presented with. 14 month olds, on the other hand, are busily learning words and re-weighing various cues, and can get distracted by the potential link between the word and the object in addition to the speech-sound change, unless the testing situation
helps facilitate their attention to the phonetic distinction or to the referential task at hand. Finally, by 17 months, infants have likely sorted out these competing cues and novel word-learning of similar sounding words can occur with ease. However, this story may be selling the 8-month-olds short. If they too are real word-learners, then a new account will need to explain why they are able to succeed in mapping a new word to a new object while 14 month olds are not.

One possibility is that what changes between 8 and 14 months is the match between the level of complexity from which the infant is best able to learn quickly. If indeed 8 month olds learn words in this task, the 1-object switch paradigm used by Stager and Werker may have found a sweet spot for 8 month olds: this paradigm may present enough audio-visual complexity to lead to learning but not so much that infants are overwhelmed, and not so little that they start to interpret the task differently, are bored, etc. This proposal resembles ideas raised by Shukla et al. (2011), namely, that infants can learn words better when the input matches their expectations. It is also kindred with ideas raised by Gogate & Hollich (2010), which suggest that mothers and infants go through iterative cycles of stability and instability in which the mother’s behavior and/or infants’ perceptual skills shift to reach the next point of stability (p. 506). Finally, this suggestion stems in part from a useful distinction, raised by many researchers, between the child’s input and the child’s intake (P. Bloom, 2002; Waxman & Lidz, 2006, inter alia).

Beyond one word, one label.

Calling the training studies described above ‘word learning’ may be a bit misleading for several reasons. First of all, it is not clear that the mechanisms that underlie word learning in these situations scale up to word learning of the native language. On the word-form side, much of the problem infants must solve has to do with isolating and recognizing words that occur somewhat sporadically over long periods of time. In training studies, a single word is stressed, occurs in isolation or at an edge, and many times in succession, greatly simplifying the word-form problem. On the conceptual side, it is often hard for infants to know what is being referred to; the referential world is far more complicated than in most word-learning studies with quick and focused in-lab training on a single word-referent pair.

In an attempt to gain traction on how infants may learn more than one word at a time from a more ambiguous referential world, L. B. Smith and Yu (2008) exposed 12 and 14 month old infants to a series of trials in which two novel objects were presented and two nonce words were heard, such that on any given trial the input was ambiguous. Both age groups in this study succeeded in learning six novel words at above-chance rates, suggesting that even with just a brief exposure to ambiguous input infants can learn a handful of word-object links.

---

7 For more on how such a ‘Goldilocks’ framework may apply in other developmental paradigms, see Kidd, Piantadosi, and Aslin (2012).
There is debate about these findings, centered on the mechanism by which this learning takes place. Some researchers believe infants could learn which word went with which object by simply tracking word-object pairings across trials, using cross-situational statistics (L. B. Smith & Yu, 2008; Yu & Smith, 2007; Yurovsky, Smith, & Yu, 2013). Others (e.g. Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell, Medina, Hafri, & Gleitman, 2013) suggest that this pattern is better explained by a mechanism in which infants form a guess on every trial, and retain this guess unless it is disconfirmed, dubbed ‘propose-but-verify.’ This proposed mechanism stems from evidence from adult and preschooler performance in a paradigm where participants are shown short video vignettes from a parent-infant video corpus, in which a key noun is replaced by a beep, and are asked to guess what word the beep represents. Participants’ pattern of guesses is consistent with the propose-but-verify mechanism, and moreover, these authors argue, this mechanism can also accounts for the Yu and Smith findings (Medina, et al., 2011; Trueswell, et al., 2013). While the debate between the cross-situational statistics and propose-but-verify mechanisms continues, the studies these mechanisms stem from show the viability of infants learning multiple words in parallel in a complicated and visually busy referential world.

Another potential concern with training studies is that they are only one part of the word-mapping problem: acquiring the link between the word and its referent. Retention of this link is rarely queried, and when it is, shows that even two year olds have trouble retaining a few newly taught words after a short delay (Horst & Samuelson, 2008; Kucker & Samuelson, 2012). This is an important area for further research. Word comprehension studies of words from the child’s native language are free from this problem because they test the retention directly, with the infants’ life history serving as the training.

Finally, in both their rhetoric and methodology, many training studies conflate the learning of proper and common nouns. Learning proper nouns is, in some ways, a much easier process. While many of the same problems we saw above in how a word is picked out still apply, on the conceptual side, infants need only to link the found word-form with a single, specific entity, usually a person. Indeed, some of the earliest learned words are such proper nouns: the infants’ own name and the word for ‘mom’ and ‘dad’ that their family uses (Mandel, et al., 1995; Tincoff & Jusczyk, 1999). In contrast, when infants learn a common noun, they need to be able to link the found word-form with an entire category of objects, of which they have only been exposed to a subset. Indeed, this is a way in which not only do common nouns differ from proper nouns, but from facts about objects (e.g. who gave you a toy) as well (cf. Markson & Bloom, 2001; Waxman & Booth, 2000). Thus, the literature on categorization and labeling is highly relevant to understanding how infants learn the meanings of words.

**Words and categories.**
By 10-14 months, infants look longer at objects during a free play period if they had recently been labeled and pointed at, as opposed to just pointed at (Baldwin & Markman, 1989), suggesting that there is something special that the act of labeling adds to direct infants’ attention to objects in the world. Building on this, Waxman and
colleagues have done a series of studies looking at how words influence infants’ category formation. For instance, Fulkerson and Waxman (2007) find that if shown a series of category members accompanied by a novel word (e.g. a series of green dinosaurs each labeled ‘toma’), 6-month-olds show a preference for a member of a new category, e.g. a green fish over a new instance of the same category, e.g. a new green dinosaur. This pattern does not apply if the category members were paired with a series of tones instead of words.

Words also help infants individuate objects. As discussed in the concepts section above, by about 4 months of age infants have a sortal for objects and by 10 months, one for people (summarized in Xu, 2007). Yet, 10 month olds surprisingly fail to expect two objects (e.g. a duck and a ball) behind an occluder if they’ve seen them appear one at a time out from behind it, while 12-month-olds succeed. However, linguistic labels interact with the individuation process. For example, 9 month olds succeed in the ball-duck version of the two-or-one task, if, while they see the objects emerge from behind the occluder, they are labeled as ‘ball’ and ‘duck,’ respectively (Xu, 2002). If they are labeled just ‘toy,’ or linked to two non-word sounds, infants fail to expect two objects, suggesting that they expect two distinct labels to refer to two distinct kinds, even though they are able to represent ‘objects’ as such in non-linguistic settings earlier on.

While the labeling study just described suggests a powerful role for labels, it leaves open whether it is labels as such, or whether children’s knowledge of the specific words facilitates their success in this task. Further evidence suggests that both of these factors matter. 10-11 month olds succeed in the no-label version of the two-or-one task if their parents report that they are familiar with the objects and their labels (Rivera & Zawaydeh, 2007; Xu & Carey, 1996), suggesting that knowing about the category makes it easier for infants to individuate at the right level of abstraction. Complementary evidence with 12 month olds shows that hearing varied labels across a set of objects (e.g. toy animals) does not lead to category formation, while hearing a consistent label does, even if these labels are all novel words to the infants (Waxman & Braun, 2005).

Results with 9 and 12 month olds demonstrate that even when the infants don’t know the labels or objects in the task, hearing two distinct labels leads them to search for or expect two objects, while hearing the same label repeated does not have this effect (Dewar & Xu, 2009). In a different paradigm, Plunkett et al. (2008) offer converging evidence: they find that 10 month old infants’ categorization of objects with correlated visual features shifts based on the nature of the link between the label and the visual features.

Thus, labels as such and labels whose referents the infants are reported to know both help them keep track of how many distinct objects they have seen, and to form categories over these objects. These studies, taken together, argue for two different, mutually compatible mechanisms: an abstract one, whereby distinct labels are expected to link to distinct count nouns, even if nothing else about the objects or labels is known; and a more concrete one, whereby seeing objects for which infants have already made the mapping makes the objects easier to track. This may be one of the many ways that language is redundant: just as we saw many cues for how to segment words, so may it be
the case that categorization and individuation of objects is facilitated by more than one kind of link between words and objects.

Thus, having seen that the link between individuation and language is a strong and multi-faceted one, we can turn to research examining the development of infants’ word-to-category mapping, and subsequently, word-to-property mapping.

**Learning words for categories and properties.**

A series of studies by Waxman and Markow (1995) helps elucidate the unfolding precision and complexity in infants’ categorization. In this work, 13 and 14 month old infants were exposed to a series of objects using either basic-level (e.g. car, plane, cow, dinosaur) or super-ordinate labels (e.g., vehicle, animal), or no labels (e.g. ‘look at this!’). At the basic level, infants in both the label and the no label condition formed categories appropriately, but at the super-ordinate level, only infants in the label condition formed the category, suggesting that especially when there are fewer visible features to form categories over, words help infants form such categories.

Further studies confirmed that this did not occur if the categories were arbitrary, and that infants’ early categorization is based on shared object class rather than property (Waxman & Markow, 1995). This leads the authors to suggest that for the early word-learner, at around 11-13 months, words are an ‘invitation to form categories’ and that children initially assume that categories are based on objecthood, i.e. correspond to common nouns in the adult vocabulary, as opposed to properties, which correspond to adjectives. Further work has shown that infants start to understand some aspects of word-to-property mapping (i.e. adjectives) around 14 months, and really sharpen this knowledge around 18 months (Booth & Waxman, 2009).

**Starting with nouns.**

By now it is clear that the literature on infant word learning focuses on nouns early on, while demonstrations of understanding of adjectives and other parts of speech occur later. While studies of infants’ first words show an early lexicon that spans many adult lexical classes (Benedict, 1979; L. Bloom, 1993) there have been far fewer studies querying infants’ understanding of words that are not object labels. Why should this be the case? It could be because nouns are learned first by infants and the literature reflects this. In contrast, it could be because nouns are learned first in the languages in which most developmental studies occur, e.g. English, which tend to be ‘noun-friendly’ languages. A final alternative, compatible with either of the preceding two, is that it is easier to test infants’ understanding of nouns than their understanding of other parts of speech, especially early on in development when indications of comprehension are harder to assess, and thus there is a research bias towards nouns because the research can utilize

---

8 A great deal has been discovered about how children learn verbs as a true syntactic category, though since this literature uniformly concerns infants 18 months and beyond, it is not included in this discussion (see Gleitman, 1990 for a thorough theoretical grounding; Waxman & Lidz, 2006 for a great overview of the critical issues and literature).
a simpler methodology. It is hard to find data to support or detract from this last alternative, though the wide range of open-class words found in early vocabularies in observational studies suggests that this alternative may have some merit; that said, these studies too find that most early words are nouns (e.g. Benedict, 1979).

**Nouns first universally.**

Among those who believe that nouns are learned first universally, there are two closely related camps. Some researchers believe that nouns are learned first because the word-object links that underlie them are more easily observable from the world, in contrast to verbs and closed class words (Gillette, Gleitman, Gleitman, & Lederer, 1999; Gleitman, 1990; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005). As Fisher and Gleitman put it, “The true beginner can only try to observe elements in the world that systematically occur with the use of particular words. This leads to success in those cases in which the word’s meaning is concrete enough to be readily observable in the flow of events: mostly nouns, but also a heterogeneous set of other words” (Fisher & Gleitman, 2002, p. 475). From here, then, word learning can jump off “because nouns typically have fewer linguistic prerequisites, they can be learned first, and can then be used as a foothold for the subsequent acquisition of words from other grammatical categories” (Waxman & Lidz, 2006, p. 126).

The suggestion from this position is that infants break into the word-meaning system using a few basic nouns, and perhaps a hodge-podge of other words that are easily observed in the input, and from there can begin to learn properties, verbs, and other parts of speech that are more fundamentally tied to argument structure, i.e. have nouns in semantic and syntactic roles, and that are more difficult to observe in the world as a ‘naïve’ observer (cf. "flee" vs. "catch" Gleitman, January, Nappa, & Trueswell, 2007).

In a not-so-distant theoretical framework, other researchers think nouns are learned first because of the way languages carve up the world into parts of speech: nouns are cross-linguistically universal (unlike, e.g. determiners), and on the concrete, easily grasped end of the conceptual spectrum (Gentner, 1982, 2006; Gentner & Boroditsky, 2001). Thus, while languages vary in terms of how verbs, prepositions, classifiers, and determiners are used (or not used), nouns are more cross-linguistically consistent. This in turn may suggest that the mechanisms used for learning nouns are more readily available, or more basically derived by the infant than those for later-learned parts of speech, for which syntactic support is integral. It is worth pointing out, however, that even when languages do vary, infants seem prepared, from the conceptual end, to learn whatever differences a language might make. For instance, Korean makes a distinction between loose and tight fit that English does not, but even English-exposed infants, at 5 months, are able to detect this distinction (Hespos & Spelke, 2004).

These two views have much in common, and both groups agree on the fundamental point that nouns are, by and large, acquired before other parts of speech by infants.9

---

9 See Macnamara, (1982, Chapter 8), for an interesting discussion of this topic.
Nouns first only in noun-friendly languages.

Other researchers have argued that there is only a noun-bias in noun-friendly languages, like English, and not in verb-friendly languages like Chinese and Korean. In Korean and Chinese, nouns can be dropped from the surface structure, leaving more bare verbs in the input to the child, and relatively fewer nouns, leading children to learn more verbs early on (Tardif, et al., 2008; Tardif, Shatz, & Naigles, 1997) However, this work is based mostly on evidence from 3-5 year olds, and both the results and their linguistic-theory underpinnings have been debated (see Waxman, et al., in press for a review). Recent evidence from 20 month olds in verb-learning and noun-learning studies in both noun- and verb-friendly languages suggests that indeed, nouns are universally relatively easily acquired, while the pattern of learning for verbs is a mixed one (Sudha Arunachalam & Waxman, 2011; Leddon, et al., 2011; Waxman, et al., in press).

Thus, the data from early word-learning suggest that word learning begins with concrete nouns, whether due to facets of the environment, the nature of linguistic input, or both. Cross-linguistic differences concerning other syntactic construction, e.g. whether a language is a ‘path’ or ‘manner’ language, are not particularly at stake in this claim, given that verb-learning as a syntactic process does not begin in earnest until further along in the second year of life (Waxman & Lidz, 2006).

Word Comprehension Theories and Summary

While I will return in the concluding chapter of this dissertation to some of the broader questions about word-learning mechanisms, I would like to conclude this chapter by sketching out a few of the existing theoretical positions about how infants learn words, many of which were mentioned above. This in turn has implications for when infants begin word learning, and which words they learn first.

As we already discussed in some detail, nouns are almost unanimously seen as the best candidate for early words. This may be, as P. Bloom has suggested, because “the most phonologically salient part of the phrase is the noun and the most semantically salient part of the context is the object” (P. Bloom, 2002, p. 117). What makes objects so salient may be, as Macnamara has suggested, that not all interpretations of acts of referring are available to the early word learner, and the names for objects are some of the first interpretations that are (Macnamara, 1982, p. 170). Both of these suggestions are in line with those we saw from Gleitman, Gentner, Waxman, and others in the preceding paragraphs.

A wholly different mechanism is proposed by those who believe young infants have a smaller cognitive endowment, either broadly, or specifically in the domain of language (L. Bloom, 1993; Plunkett, Sinha, Muller, & Strandsby, 1992; Sloutsky, 2003; L. B. Smith & Yu, 2008; Yu & Smith, 2007, inter alia). These researchers suggest that infants learn nouns first because of the simple statistics in the world they experience, using ‘dumb attentional mechanisms’ (L. B. Smith, Jones, & Landau, 1996). For instance, under this view, infants learn object names because they see the object and hear it named across multiple situations over time, with no need for infants’ attribution of intent, belief, or cause to the visual situation to come into play (e.g. L. B. Smith et al.,
The view expressed by Gogate in the work reviewed above is compatible with this school of thought: namely, in her analysis that learning the relation ‘goes with’ is the necessary prerequisite for learning ‘stands for’ (Gogate, 2001).

This hypothesis perhaps provides a more straightforward and parsimonious account than those positing more mental structure within the infant that is recruited under various conditions, timelines, and with varying degrees of success; some version of this hypothesis may turn out to explain, for instance, how the shape bias develops (L. B. Smith, 2000, 2005). However, while there are data that are compatible with this hypothesis (but also others), e.g. Yu & Smith (2007), it has also been criticized for its ability to account for various aspects of word learning. For instance, P. Bloom lays out a series of concerns with such a word-learning mechanism. First, it doesn’t account for the proportion of the time that infants are hearing object labels for non-present objects, which is perhaps as high as 30% (Harris, Jones, & Grant, 1983), nor are children found to make the frequent mapping errors that would be predicted given anything close to such a rate of referent-less object naming. Second, the theory doesn’t explain how abstract words or words for imaginary entities would be learned. Finally, given how good nonhuman primates are at learning statistics of their environment, this mechanism doesn’t explain why they fail to learn words, leading P. Bloom to conclude that “statistical co-variation between word and perceiv is neither necessary nor sufficient for word learning” (P. Bloom, 2002, p. 59). This seems a pretty damning set of criticisms.

A contrasting account to the statistical learning proposal, mentioned briefly above, is that offered by Waxman and Gelman (2009). These authors suggest that it is not correct to characterize the word learning infant as a ‘data analyst’ but rather that she should be considered a ‘theorist’ as well. They offer evidence suggesting that words provide abstract reference above and beyond associations to the object, that words are not only based on visual features alone, but rather that infants do invoke knowledge about causes, intent, and animacy in learning words; that infants’ sensitivity to different kinds of word-referent links is not unitary but shifts over time; and finally that words cannot be considered in a vacuum but must be considered in the linguistic and social midst in which they occur. The evidence these authors draw on is wide ranging and seems too to leave the statistical learning account an implausible one.

The dissertation that follows, much like the theories just outlined, can only offer possibilities for how infants learn words indirectly, by measuring when they understand words of various kinds. From there, the current theories, if found wanting, can be reassessed. The preceding decades of research seem to suggest that few if any common nouns are learned before infants’ first birthday. By then, infants’ native language phonology has solidified, though word-segmentation and word-form recognition have already been active in the child’s parsing of the linguistic input from around 6-9 months onward. The literature also suggests that when word-learning does begin, it likely begins with nouns. Moreover, the nouns infants may learn first, based on their conceptual knowledge, are suggested to be salient, stand-alone objects that adhere to the various criteria established by Spelke and colleagues, with regard to boundedness, connectedness, coherence, etc. (Spelke, et al., 1992; Spelke, et al., 1995).
Having set the stage with what we know already, and what questions remain, I will argue for three novel contributions to the literature, based on the experiments described below. First, when infants learn words, some of the earliest words they learn are, perhaps unexpectedly, foods and body-parts (Chapter 2). Second, soon after we first see signs of infants’ comprehension of these early concrete nouns, infants demonstrate comprehension of a heterogeneous group of words ranging from greetings to verbs to exclamations; the ability to learn this broad range of word meanings suggests impressive abstraction is taking place (Chapter 3). Third, right from the beginnings of word comprehension, infants make good assumptions about what sort of acoustic differences matter when comprehending words, and what differences do not (Chapter 4). Finally, I will conclude by summarizing what I’ve shown, situating it among word-learning theories and timelines, and discussing the logical next steps for this program of research (Chapter 5).
Chapter 2: At Six to Nine Months, Human Infants Know the Meanings of Many Common Nouns

This work was originally published in the Proceedings of the National Academy of Sciences:


ABSTRACT

It is widely accepted that infants begin learning their native language not by learning words, but by discovering features of the speech signal: consonants, vowels, and combinations of these sounds. Learning to understand words, as opposed to just perceiving their sounds, is said to come later, between 9 and 15 months of age, when infants develop a capacity for interpreting others’ goals and intentions. Here, we demonstrate that this consensus about the developmental sequence of human language learning is flawed: in fact, infants already know the meanings of several common words from 6 months onward. We presented 6- to 9-month-old infants with sets of pictures to view while their parent named a picture in each set. Over this entire age range, infants directed their gaze to the named pictures, indicating their understanding of spoken words. Because the words were not trained in the laboratory, the results show for the first time that even young infants learn ordinary words through daily experience with language. This surprising accomplishment indicates that contrary to prevailing beliefs, either infants can already grasp the referential intentions of adults at 6 months, or infants can learn words before this ability emerges. The precocious discovery of word meanings suggests a new perspective in which learning vocabulary and learning the sound structure of spoken language go hand in hand as language acquisition begins.

Introduction

Most children do not say their first words until around their first birthday. Nonetheless, infants know some aspects of their language’s sound structure by 6–12 months: they learn to perceive their native language’s consonant and vowel categories (1-4), they recognize the auditory form of frequent words (5, 6), and they use these stored word forms to draw generalizations about the sound patterns of their language (7, 8), using cognitive capacities for pattern-finding (9, 10). Although this learning about regularities in speech reveals impressive perceptual and analytical skill, it is generally accepted that young infants do not know the meanings of common words. Indeed, while some experimental work has shown that young infants can associate syllables with individual objects after laboratory training (11), prior experimental tests have failed to detect understanding of common native-language words before around 12 months (12).

Infants are, on the whole, proficient and precocious learners in other domains (13), so why would learning word meanings be difficult for them? The most prominent
hypothesis is that true word learning is possible only when infants can grasp speakers’ referential intentions and understand language as a motivated, communicative activity (14-17). Evidence that infants only begin to understand other humans as intentional agents at around 9 to 10.5 months has been argued to explain the earliest emergence of word learning shortly thereafter (17). Understanding reference is said to be necessary for word learning because the natural conditions of language use do not support the simple associations that underlie, for example, trained dogs’ ability to fetch specific toys on command (18). The statistical connection between instances of words and the details of infants’ observations is tenuous: parents do not reliably say “doll” in the exclusive presence of dolls, and they say “Hi, I’m home!” more often than “Daddy is moving through the doorway!” (19). Furthermore, words (excepting proper names) refer to categories, not individuals, and the learner must discover each category and its boundaries. Thus, while infants can link ‘mommy’ with films of their mother, these labels do not indicate that infants have induced the relevant category (20). Because of these complexities inherent in language understanding, the predominant view is that word learning is possible only when children can surmise others’ intentions enough to constrain the infinite range of possible word meanings, a skill believed to develop gradually after 9 months (17). Until that age, infants’ native language learning is held to be restricted to speech signal analysis (21).

In the present research, we examined young infants’ knowledge of word meaning using a new variant of a task called “language-guided looking” or “looking-while-listening” (22, 23). In this method, infants’ fixations to named pictures are used to measure word understanding. Infants are presented with visual displays, usually of two discrete images, one of which is labeled in a spoken sentence like “Look at the apple.” (24, 25). In our variant, the parent uttered each sentence, prompted over headphones with a pre-recorded sentence, ensuring that infants (n=33) heard the words pronounced by the familiar voice of their parent. Each infant saw two kinds of trial: trials with 2 discrete images (paired-picture trials) and trials with a single complex scene (scene trials) (see Methods, Fig.1, Fig. S1, and Table 1).

Two word categories were tested: food-related words and body-part words. Paired-picture trials (n=32) presented one image from each category (e.g. apple – mouth), and scene trials presented one image (n=16) depicting several category members together (e.g. a full-length picture of a boy, a close-up of a face, or a table with food-related items on it). All pairs and scenes occurred in multiple instantiations within and between infants (e.g., there were 2 different ‘apple’ photos, and two different ‘full-body’ photos; see Fig. S1).

Results

Children who understood a word were expected to fixate the target picture more upon hearing it named. To evaluate this, the two trial types were analyzed separately, because their demands are distinct and the ideal analytical methods are different, particularly in how to best correct for infants’ preferences for individual pictures. (An analysis of both trial types using the same dependent measure is given in the Supplementary Information (SI) and in Table S1.)
For both analyses, the post-target analysis window extended from 367ms to 3500 ms after the onset of the spoken target word (see Fig S2). The 367-ms starting time is the standard in the field and allows for the time required to initiate an eye movement in response to the speech signal; earlier fixation responses are unlikely to be reactions to the signal. The 3500-ms window offset is later than the 2000-ms offset that is typically employed with older children. It was implemented here because in previous research testing the 12-24 month age range, children have been discovered to get faster with increasing age and experience with words; thus we assumed that younger children would require more time to demonstrate recognition.

<table>
<thead>
<tr>
<th>Body Items</th>
<th>ear, eyes, face, foot, feet, hair, hand, hands, leg, legs, mouth, nose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food Items</td>
<td>apple, banana, bottle, cookie, juice, milk, spoon, yogurt</td>
</tr>
<tr>
<td>Scene Body Pictures</td>
<td><em>Face</em>: eyes, hair, ear, mouth</td>
</tr>
<tr>
<td>Scene Food Pictures</td>
<td><em>Body</em>: face, legs, hands, feet</td>
</tr>
<tr>
<td>Scene Food Pictures</td>
<td><em>Tabletop1</em>: milk, juice, spoon, banana</td>
</tr>
<tr>
<td>Scene Food Pictures</td>
<td><em>Tabletop2</em>: apple, cookie, yogurt, bottle</td>
</tr>
<tr>
<td>Paired-Pictures</td>
<td>apple-mouth, banana-hair, bottle-leg, cookie-eyes, juice-nose, milk-foot, spoon ear, yogurt-hand</td>
</tr>
</tbody>
</table>

**Table 1.** Target items. The top half of the table lists all food and body part target words that were tested. The lower half of the table lists the target words for each trial type, and indicates the image (scene trials) or yoked pair (paired-picture trials) within which each target word’s referent picture occurred.

**Figure 1.** Experimental Setup. The child sat on her parent’s lap and was presented with images and sounds from a computer equipped with an eye tracker and speakers. The experimenter sat behind a screen, and
was not visible to the infant. The experimenter controlled presentation of stimuli, and monitored the child on a live-feed camera. A backup video recording of the session was made to allow for confirmation of the validity of gross characteristics of the eyetracking data stream. The figure shows an example of images presented on a paired-picture trial testing ‘banana’ or ‘hair.’

For paired-picture trials, word recognition performance was operationalized as a difference of fixation proportions: for paired pictures A and B, the fixation to picture A relative to B when A was the target, minus fixation to A when A was the distracter.* For example, given the pair of images hair/banana, a child’s performance was given as how much she looked at ‘hair’ when it was named as the target, relative to her looking at ‘hair’ when ‘banana’ was the named target. Positive difference scores are consistent with word understanding. This pair-based analysis corrects for infants’ picture preferences without relying on infants’ looking during the portion of the trial before the mother speaks. 26 of the 33 6–9 month olds (M=7.44 months, SD=1.26) showed a positive mean difference score (all 33 subjects’ M=0.074, P=.0005, Wilcoxon Test; P=.001, binomial test). Children showed positive performance on 6 of 8 item-pairs (M=.065, P=.020, Wilcoxon Test). Figure 2 illustrates these results showing the 6–7 month olds and 8–9 month olds separately.

For scene trials, word recognition performance was operationalized as the proportion of target looking upon hearing the target word (367-3500ms post target onset), minus the proportion of target looking before hearing the word (from when pictures were displayed until just before target onset)(see Fig. S2). This analysis corrects for fixation preferences within portions of the scene, preventing an advantage for targets that take up more of the scene. 22 of 33 infants showed a positive proportion of target looking; performance was statistically significant over subject means (M=.042, P=.020, Wilcoxon Test; Fig. 2). Infants showed positive performance on 12 of 16 items; performance over item means fell short of significance (M=.023, P=.058, Wilcoxon Test; Fig. 2).

Infants of 8–9 months are known to be capable of learning the sound forms of words and retaining them over long intervals (6), whereas infants of 6-7 months have thus far shown much more limited word-form knowledge (e.g., 26). It is therefore of interest to determine whether the present findings are due only to the older children in the sample. This was not the case. Considering the subject-means data in Figure 2 it is clear that for both types of trial, at both ages most children scored above zero. On the paired trials, performance was significantly above chance levels in each age group (6-7, M=.058, P=.027; 8-9, M=.082, P=.0052). In the scene trials, evidence of recognition was strong in 6-7 month olds (M=.068, P=.015) but less strong in 8-9 month olds (M=.013, P=.27) though these age groups were not significantly different from one another (paired-picture trials: M=.036, P=.37; scene trials: M=.067, P=.093). The apparently inferior performance of the 8–9 month olds on the scene trials may be traced to their tendency to fixate the ‘eyes’ and ‘face’ regions before the mother named any pictures (see figure 2, part H). This tendency, which may have its origins in previously observed developmental changes in infants’ attention to social stimuli (27), did not interfere with infants’ demonstrating recognition in the paired picture context, but impeded accurate measurement of 8-9 month olds’ word recognition in the scenes containing faces.
A correlational analysis over the 6–9 month range indicated that performance on paired-picture trials was not correlated with age ($\tau=.042, P=.75$). Performance on scene trials was negatively correlated with age ($\tau=-.25, P=.039$); however, excluding the two words ‘eyes’ and ‘face’ (or just ‘eyes’ or just ‘face’) the correlation of performance with age was negligible ($\tau=0.015, P=.91$). The lack of a positive correlation with age, and the consistently strong performance of the 6–7 month olds, confirm that the word recognition performance of the 6–9 month old sample cannot be attributed to the older children alone.

The timecourse of infants’ picture fixation is shown in Figure 3. The Figure presents data from the 33 6–9-month olds, as well as results from three older groups of children tested in the same procedure (see SI Text S1). Children initially fixated the target and distracter equally (averaging over items); then, upon hearing the target word, they shifted gaze to the named picture, thenceforth remaining above chance levels of target-looking over most of the trial. While most infants showed knowledge of the meanings of most items, target fixation performance at 6–9 months and even 10–13 months was below levels shown by slightly older children (Fig. 3). The data suggest a discontinuity in performance at around 14 months: performance was stable with respect to age before 14 months, and was substantially better afterward. We speculate that this phenomenon reflects the acquisition of linguistic knowledge and the development of social or other communicative skills, a topic we return to in the discussion. A more detailed analysis of the developmental pattern of results is given in the Supplementary Information (Text S1).

Two additional measures of 6–9 month old infants’ word knowledge were obtained from their parents: the MacArthur-Bates Communicative Development Inventory (CDI), which is a vocabulary checklist originally intended for children 8 months and older (28), and an item exposure survey asking parents to estimate how often their child heard our target items, on a scale from ‘never’ to ‘several times a day’. The modal response from parents on the CDI was that their child did not know any of the 395 words on the inventory; further, no parent reported that his or her child was producing any of the words tested in our experiment (see Table 2). Parental ratings of item exposure did not correlate significantly with scene or paired-picture trial performance ($\tau=-.022, P=.65$ (scene); $\tau=.0079, P=.83$ (paired-picture), Kendall’s correlation test). Thus, there was no indication that the knowledge that infants revealed in the experiment was apparent to the infants’ parents.

**Discussion**

The present findings provide an important contribution to our understanding of language acquisition, showing that by 6-9 months, infants have already begun to link words with their referents, over a range of food and body-part terms. The two trial types indicated complementary abilities. Success on paired-picture trials showed that infants could understand words whose referents were presented in extremely stripped-down contexts: e.g., a nose without the eyes or mouth. By contrast, performance on the scene trials showed that infants could differentiate at least some of the tested words from
Figure 2. Subject and Item/Pair Means for 6-7 and 8-9 month olds. All data (A:H) calculated over a window from 367-3500 ms post-target word onset. Subject mean difference scores are shown for paired-picture trials for 6-7 month olds (A) and 8-9 month olds (B). Subject mean increases in target looking, corrected for baseline looking, are displayed for scene trials for 6-7 month olds (C) and 8-9 month olds (D). Item-Pair mean difference scores are shown for paired-picture trials for 6-7 month olds (E) and 8-9 month olds (F). Item mean increases in target looking, corrected for baseline looking, are given for scene trials for 6-7 month olds (G) and 8-9 month olds (H). (E:H) error bars represent bootstrapped nonparametric 95% confidence intervals. On the right of each subplot is a histogram of the responses in the main plot; all histograms show more positive than negative responses for each subset of subjects and of items/pairs.

Figure 3: Timecourse of infants’ picture fixation on paired-picture trials, averaged over infants in four age groups. The ordinate shows the mean proportion of infants who were looking at the named (target) picture at each moment in time. Error bars indicate standard errors of the mean, with means computed over subjects in each age range. At all four ages, target fixation rose from about 0.50 (chance) shortly after the onset of the spoken word. Overall, accuracy in fixating the named picture increased with age across the age groups. See Text S1 for further details.

Several features of these results merit further exploration. The present study showed that 6–7 month olds understand something of the meaning of at least some words for foods and body parts. The results do not establish the size of infants’ vocabularies, nor do they prove exactly which of the tested words each child knew. The item-wise histograms in Figure 2 show that mean performance was greater than zero, and that there

related alternatives: for example, an infant who heard ‘banana’ and then looked at the banana in a tabletop scene containing several objects from typical meal-time contexts provided evidence of distinguishing the ‘banana’ from semantically related objects. On both trial-types, infants demonstrated abstraction from their experience, in that the pictures we selected were not adapted to children’s individual experiences in any way. Each word was tested on three trials (twice in picture pairs and once in a scene). In each case the image was different and the spoken words, being produced ‘live’ by the parent, were never exactly the same. Infants therefore showed generalization in the way that language normally demands: common nouns refer to categories, not a specific instance, and spoken words are consistent phonologically (in their sequence of consonants and vowels) but not acoustically (e.g. in the details of each instance’s pitch and duration).
was roughly normally-distributed variation around that mean. Some variation is
undoubtedly due to chance, but it is also quite likely that there were some words that
some children did not know. After all, children’s experience is various, and the items
tested were not calibrated to each child’s history. Confirming knowledge of particular
words would require a design devoting more test trials to a smaller number of different
words. The fact that the present results are statistically significant in analytical models
that include subjects and items as random factors (see Text S1 and Table S1) ensures that
the conclusions are not based on just a few words or a few children; however, the present
design does not support strong conclusions about individual words or individual children.

We have focused here on the youngest children in the sample, because the
assumption that young infants learn about speech sounds but not about word meaning has
predominated for at least 25 years. However, we also found a surprising developmental
pattern when considering a much wider age range, from 6 to 20 months. Previous studies
have not explored this age range using consistent materials and procedures. Doing so
allowed us to document little change in performance from 6 to 13 months, and then a
substantial improvement at around 14 months, with some gains observed thereafter. This
pattern raises two questions: why was there no apparent improvement in the youngest
half of the sample, and what caused the change in performance at around 14 months?

One explanation of the seeming lack of developmental change from 6 to 13
months is that it is artifactual. First, informally speaking infants of about 9–12 months
are often more difficult to evaluate in experimental procedures than younger or older
children, seeming more distractible and more likely to become fussy. This might
be a product of infants’ eagerness to exercise their rapidly changing motor skills and a
consequent lack of attention to the experimental materials, thus masking underlying
developments in linguistic knowledge or ability. Of course, we might also speculate that
a lack of attention to language among some 8–12 month olds who are prioritizing some
other cognitive domain indicates something true about their mental life outside the
laboratory as well (e.g., 29).

An alternative account for the apparent lack of improvement in performance from
6 to 9 months is tied to an explanation for the material elevation in performance around
14 months. We speculate that younger and older children might learn words, interpret
sentences, and conceptualize the experimental situation in quite different ways. In the
domain of word learning, infants may be restricted to relatively inefficient learning
strategies, a point we return to below. Infants are also likely to be much less
sophisticated in their interpretation of the sentences in which our target words were
embedded. Knowledge of English sentences and English syntax make a sentence like
“Can you find the juice?” interpretable, not just as a string of syllables followed by a
familiar word, but as a hierarchy of syntactic phrases unfolding in somewhat predictable
ways. Understanding the sentence is likely to make the target words easier to grasp. It is
also possible that at around 14 months many infants begin to catch on to the nature of the
experiment, regarding it as a repetitive game of object-searching, and that this helps them
focus their attention on the task. These factors could explain the apparent lack of change
in performance from 6 to 9 months. The relatively low levels of target looking at the
younger ages may have been feasible with basic knowledge of word meaning but not a
richer understanding of the words, a better grasp of the sentences, or a helpful conceptualization of the task. Of course, these comments are speculative: while developmental improvements along these lines are to be expected at a gross level, the present experiment was not intended to evaluate these possibilities.

How do 6–7 month olds learn words? The key finding of this study is that infants recognized words and demonstrated through their behavior that they knew something of the meaning of those words. Because the study involved no training, the result implies that 6-7 month olds learned the words through their daily experience. The predominant account of word learning holds that intention-reading is a fundamental prerequisite. One interpretation of our results is that the relevant social-cognitive skills are available earlier in development than previously assumed. This is consistent with recent evidence of young infants’ early sophistication in social cognition (30, 31). For example, 6-month-olds follow an adult’s gaze when the adult signals an intent to communicate, but not otherwise (32). Infants’ appreciation of their parents’ referential intentions, though perhaps limited at 6 months, may narrow the range of potential word referents, making word learning possible.

Some theorists argue, by contrast, that domain-general cognitive capacities suffice for word learning, without invoking any understanding of the referential nature of words or of others’ intentional states (33, 34). This association-based kind of account would explain our results by appealing to infants’ ability to track consistent features of their physical environment when hearing words, progressively hypothesis-testing (with or without an active intent to learn) until the referents of the words were isolated. In considering this hypothesis we note that many of our target words did not refer to distinct, bounded objects, which have been suggested as good defaults for such hypotheses (35). Infants here performed well across an array of items containing well-delineated objects (e.g. cookie, bottle); amorphous substances (e.g. milk, juice); and unclearly bounded body parts (e.g., nose, hands) (see Fig. S1 for visual stimuli and regions of interest). This does not rule out associationist theories of word learning, but it raises the stakes; an association-based account (or indeed any account) cannot rely strongly on a bounded-object bias to limit its search space.

Our results do not imply that infants have an understanding of words that is comparable to adults’, or even to older children’s. Although infants did generalize from their experience to the particular photographs we happened to choose, their categories of each object may nevertheless be different from adults’, possibly by being based more strongly on perceptual attributes than functional ones, for example. It is also not clear that our target words, which were all grammatically nouns of English, are interpreted by infants using a linguistic system that includes Noun as a category. Finally, although there is some evidence that young infants’ knowledge of the sound forms of words is accurate (for example, infants recognize words more readily when the words are pronounced correctly than when their forms are altered (26)), the present study did not test the details of infants’ speech perception.

Our findings indicate that native-language learning in the second half of the first year goes beyond the acquisition of sound structure. The fact that even 6-7 month olds learn words suggests that conceptual and linguistic categories may influence one another
in development from the beginning (36) and that aspects of meaning are available to
guide other linguistic inferences currently thought to depend only on distributional
analysis of phonological regularities (37, 38). Understanding word meaning could also
support the acquisition of syntax by guiding infants’ inferences about how nouns and
words from other word classes are placed in sentences. Precocious word learning also
helps explain why hearing-impaired infants identified for fitting with cochlear implants
before 6 months reveal better language skills at 2 years than children identified just a few
months later: 6-month-olds who can hear are already learning words (39).

More generally, these results address one of the central mysteries of language
acquisition: how children demonstrate proficiency in their native language so rapidly,
typically speaking hundreds of words by the age of two years. Part of the solution, it
appears, is that learning begins very early in life, hidden from view; even before they
begin to babble, infants understand some of what we tell them.

<table>
<thead>
<tr>
<th>Measure (# of items)</th>
<th>Range</th>
<th>Mean</th>
<th>Median</th>
<th>Mode (n=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDI understood (n=395)</td>
<td>0 - 71</td>
<td>19</td>
<td>7.5</td>
<td>0 (n=7)</td>
</tr>
<tr>
<td>CDI said (n=395)</td>
<td>0 - 7</td>
<td>0.83</td>
<td>0</td>
<td>0 (n=20)</td>
</tr>
<tr>
<td>Test Items from CDI understood (n=16)</td>
<td>0 - 10</td>
<td>2.10</td>
<td>1</td>
<td>0 (n=13)</td>
</tr>
<tr>
<td>Test Items from CDI said (n=16)</td>
<td>0 - 0</td>
<td>0</td>
<td>0</td>
<td>0 (n=30)</td>
</tr>
</tbody>
</table>

Table 2. Summary statistics for CDI completed by 30 of our 33 caregivers

Materials and Methods

Participants.

Infants were 33 6-9 month-old infants (M=7.45 mo., R=5.99-10.00 mo., 19
female). Infants were recruited from the Philadelphia area by mail, email, phone, and in
person. All children were healthy, carried full-term, and heard 75% English or more in
the home. None had a history of chronic ear infections. A second set of 50 children
ranging in age from 10-20 month olds (M= 14.1 mo., R=10.13-20.85 mo., 27 female)
were also recruited, using the same methods and criteria, to participate in the
developmental portion of the study reported in the Text S1 and Fig. 3.

Materials.

On each trial, parents spoke a single sentence to their child. To do this, they
repeated verbatim a pre-recorded sentence they heard over headphones. These pre-
recorded sentences were produced by a native English-speaking woman, recorded in a
sound-treated room. Each sentence that was presented to parents followed one of four
different formats: “Can you find the X?”, “Where’s the X?”, “Do you see the X?”, and
“Look at the X!”, where X stands for the target word (only 1 sentence format was used
per item; see fig. S2). Sentence formats varied across trials pseudo-randomly. The
sentences were uttered at a slow speed, about 4 syllables per second, with slightly
exaggerated intonation, which parents were asked to emulate. The recorded sentences
were 1-1.5 s in duration, and were presented to parents at loudness levels of 31.5-
33.75dB. Pretesting determined that speech presented at this volume over the closed-ear
headphones was audible only to the parent.
Visual stimuli were displayed on a 34.7 cm by 26.0 cm LCD 75 dpi screen. On paired-picture trials two 16.9 cm by 12.7 cm photos were displayed on the right and left side of the screen on a grey background; side of presentation was counterbalanced across trials and trial-orders. There were 32 such photos: viz., two instances each of 16 items (see Table 1, Fig. 1, and Fig. S1). Photos were edited so that their relative size and brightness were approximately equivalent.

On scene trials one photo was displayed in the center of the screen on a gray background. The photos were of people (whole body, clothed), faces, and tabletops with four food items on them (see Table 1 and Fig. S1). There were two photos of faces (widths 21.84 cm and 25.40 cm), two photos of bodies (widths 16.08 cm and 20.93 cm) and two photos of tabletops: one with milk, juice, a spoon, and a banana on it, and another with a cookie, an apple, a bottle, and yogurt on it (all widths 34.67 cm). Infants only saw one of these images on a given scene trial.

Additionally, every eighth trial a two-second movie, featuring colorful shapes and smiley faces flitting around the screen accompanied by a whistling sound, was played to maintain infants’ interest.

**Apparatus and procedure.**

Infant visual fixation data were collected using an Eyelink CL computer (SR Research), which provides an average accuracy of 0.5°, sampling from one eye at 500Hz. It operates using an eye-tracking camera at the bottom of the computer screen; no equipment is mounted on the child’s head, except a small sticker with a high-contrast pattern on it for aiding the eyetracker in keeping the infant’s position.

Before the experiment began, the procedure was explained to parents, informed consent was obtained, and a vocabulary checklist and word-exposure survey were completed. The child and parent were then led to the dimly lit testing room, where the infant sat on his or her parent’s lap facing a computer display (see Fig. 1). Parents wore a visor that prevented them from seeing the screen, and headphones over which they were prompted with the target sentence. The prerecorded sentence was then followed by a tone that indicated to the parent that she should begin repeating the sentence she had just heard (see Fig. S2).

Infants were presented with 48 test trials in two interspersed conditions: 32 paired-picture trials, and 16 scene trials. There were eight foods and eight body part items in each condition. During paired-picture trials, infants saw two images on the screen, one from the food category and one from the body-part category. On scene trials, infants saw a single complex image of a body, face, or one of two tabletops with 4 food items on it (see Fig. S1). Thus, paired-picture trials presented targets across the domains of food and body parts whereas the scene trials presented targets within one of these domains. Images were shown for 3.5s or 4s after target onset (paired-picture and scene trials, respectively, see Fig. S2); the length of time before the parent said the target varied from trial to trial, averaging approximately 3-4s. All subjects saw both trial types, and subjects were randomly assigned to one of two pseudo-randomized trial orders, which counterbalanced side, picture instance, and ordering of images and target items. The experiment lasted
approximately 15-20 minutes, after which families were compensated with a choice of $20 or two children’s books. The entire visit lasted approximately 45 minutes.

* Because the pairs are yoked, for each pair A-B the values of this measure for A and for B are arithmetically redundant (the value for A is necessarily the complement of the value for B). Thus, the item results are presented item-pair by item-pair (see Fig. 2).

Acknowledgements
This work was funded by NSF-GRFP an NSF-IGERT grants to EB and NSF grant HSD-0433567 to Delphine Dahan and D.S., and NIH grant R01-HD049681 to DS. Thanks to students and faculty at Penn’s IRCS for useful discussions of these findings, especially Josef Fruehwald, and to Richard Aslin and Lila Gleitman for comments.

Supplementary Information: Text S1
Analyses of three older groups of children and modeling of all age groups.
In addition to 6-9-month olds, we tested three additional older groups of subjects (Fig. 3): 10-13-month-olds (n=30, M=12.13 months, SD=1.08 months; 18 female), 14-16-month-olds (n=7, M=14.49 months, SD=1.04 months; 3 female), and 18-20-month-olds (n=13, M=19.38 months, SD=0.86 months; 6 female). The purpose of testing older children was to examine the developmental course of word recognition beyond 9 months, and to confirm that the performance of older children on the paired-picture trials, in our modified version of the procedure, would accord with performance of older children in similar prior experiments.

As in the analyses for 6-9 month olds, we analyzed the data with measures that corrected for inherent preferences for the pictures (difference scores for paired trials and pre-target corrected proportions of target looking for the scene trials; see main text). Additionally, we modeled the data for each trial type using the same dependent variable for both scene and paired-picture trials for all four group of children. As in the analyses reported elsewhere, target-looking performance was evaluated over a time window from 367 to 3500 ms post target onset.

A separate hierarchical logistic regression model was created for each group of children (6-9 months, 10-13 months, 14-16 months, and 18-20 months), for each trial type (paired-picture and scene). Phase of trial (pre-target utterance vs. post-target utterance) was included as a fixed effect predictor, and subject and item were included as random effects. Each model predicts the (log of) the ratio of target to distracter looking, as calculated by counting time-bins with looks to the target and time-bins with looks to the distracter(s). The input to each model was subject X item level data. Word recognition is shown when the trial phase predictor is significant and positive. Results for the predictors in each model are found in Table S2.

The results of the models for each of our groups of children were analogous for each trial type, i.e. there were differences in effect sizes, but the patterns of significance and the overall direction of the estimates were the same. Thus, for paired-picture trials, performance in the pre-target period was not significantly different from chance, as expected, while performance in the post-target period was significantly positive. For scene trials, performance in the pre-target period was significantly below chance, that is,
as expected subjects looked at the three distracters more than at the target before anything was said (the combined area of the distracters is always larger than of a given target, see figure S1). In the post-target period, performance was significant for both trial types in all four age groups (see Table S1 for estimates and p-values; all significant p-values are <.001.) These modeling results confirm the results reported in the main text.

Turning to the dependent measures corrected for picture-fixation biases as in the main text, for paired picture trials subject means and item-pair means were significantly above chance for all of the three older age groups. In paired-picture trials, 20 out of 30 10-13-month-olds showed a positive proportion of target looking; performance was statistically significant over subject means \((M=0.055, P=0.010\) by Wilcoxon Test; all following tests are Wilcoxon tests). Subjects showed positive performance on 7 out of 8 pairs (all but eyes-cookie); performance over item-pair means was significant \((M=0.059, P=0.008)\). 7 out of 7 14-16-month-olds showed a positive proportion of target looking; performance was statistically significant over subject means \((M=0.29, P=0.008)\). Subjects showed positive performance on 8 out of 8 pairs; performance over item-pair means was significant \((M=0.28, P=0.004)\). Finally, 12 out of 13 18-20-month-olds showed a positive proportion of target looking; performance was statistically significant over subject means \((M=0.33, P=0.008)\). Subjects showed positive performance on 8 out of 8 pairs; performance over item-pair means was significant \((M=0.30, P=0.004)\).

For scene trials, overall performance of 10-13-month-olds was not above chance over subject means, though their subject means did not differ significantly from the successful performance of the 6-9 month olds taken together, nor of the 6-7 month olds alone (estimate of differences: .035 for 6-9 mo., .067 for 6-7mo., .013 for 8-9mo.; all p>0.10). 14 out of 30 10-13-month-olds showed a positive proportion of target looking considering all items \((M=-0.0002, P=0.51)\). Performance was positive for 12/16 items, marginally significant over all items, \((n=16, p=.080, m=.023)\), significant over all items excluding ‘eyes’ \((n=15, p=.047, m=.033)\), and marginally significant over all items excluding ‘face’ and ‘eyes’ \((n=14, P=.059, m=.038)\). The large impact of these two items on the data was also found in the 6-9-month olds (see Main Text); low performance for this item is in keeping with a shift in social cognitive abilities occurring around 9 months of age, after which infants are better attuned to follow eyegaze (A.N. Meltzoff & R. Brooks, 2007). It is possible, then, that the poor performance of the 10–13 month olds on these trials (as opposed to their good performance on the more traditional paired-picture trials) was related to the difficulty of measuring language-related looking using images containing faces (and eyes in particular), since pre-target utterance levels of looking at ‘face’ and ‘eyes’ are very high. The dependence of this result on a small minority of the items suggests caution in interpreting this pattern as characteristic of this age group.

Among the older children, 5 out of 7 14-16-month-olds showed a positive proportion of target looking; performance was statistically significant over subject means \((M=0.13, P=0.039)\). Subjects showed positive performance on 12 out of 16 items; performance over item means was significant \((M=0.11, P=0.022)\). Finally, 12 out of 13 18-20-month-olds showed a positive proportion of target looking; performance was statistically significant over subject means \((M=0.16, P<0.0003)\). Subjects showed
positive performance on 14 out of 16 items; performance over item means was significant ($M=0.18$, $P<0.0001$).

In sum, testing of three older groups of children revealed a developmental pattern in which the youngest infants (6-13 months) performed at similar levels, while one-year-olds (14-20 months) performed substantially better. Potential explanations for this developmental trajectory are offered in the Discussion.

**Figure S1.** Sample visual stimuli and regions of interest. The top two rows show the scene stimuli. The bottom two rows show the paired-picture stimuli. Multiple photographs were used for each target image across trials and subjects. The bottom row of each set indicates, using outlining yellow lines and yellow shading, where the regions of interest were for analyses. These lines and shading were not visible during the study. For paired-picture trials every instance of every image appeared on the left and on the right across trials and subjects.

**Figure S2.** Experimental Timeline: sequence of one test trial. Parent and child heard a beep as the pictures appeared (musical note symbol). Then the parent heard the target sentence over headphones; both parent and child heard the click sound (percussion note symbol); and the parent uttered the target sentence. At the moment the parent began to say the target word, the experimenter started a timer. The pictures remained on the screen for 3.5 or 4s after this point for paired and scene trials, respectively. Exact timing varied from trial to trial, but the click was played 1-1.5s after the trial onset, and parent said the target item afterward. Each trial lasted about 7.5s.
Table S1. Model coefficients, variance, and significance estimates in hierarchical logistic models of looking results. For both trial types (scene and paired-picture) the dependent variable was the logarithm of the ratio of target to distracter looking, as measured by summing the number of 20-ms time-frames in which infants looked at the target or the distracter(s). Ratios were computed for each item within each subject. Significant negative ‘intercept’ values indicate greater looking at distracters than targets in the portion of the trial before the target word was spoken (an expected result on scene trials, which had three distracters for each target). Significant positive ‘phase of trial’ values indicate greater looking at the target after the parent said the target word than before. Random effects estimates for subjects and items (not shown) were included in all models.
Chapter 3: The Acquisition of Abstract Words by Young Infants


ABSTRACT

Young infants’ learning of words for abstract concepts like ‘all gone’ and ‘eat,’ in contrast to their learning of more concrete words like ‘apple’ and ‘shoe,’ may follow a relatively protracted developmental course. We examined whether infants know such abstract words. Parents named one of two events shown in side-by-side videos while their 6-16-month-old infants (n=98) watched. On average, infants successfully looked at the named video by 10 months, but not earlier, and infants’ looking at the named referent increased robustly at around 14 months. 6-month-olds already understand concrete words in this task (Bergelson & Swingley, 2012). A video-corpus analysis of unscripted mother-infant interaction showed that mothers used the tested abstract words less often in the presence of their referent events than they used concrete words in the presence of their referent objects. We suggest that referential uncertainty in abstract words’ teaching conditions may explain the later acquisition of abstract than concrete words, and we discuss the possible role of changes in social-cognitive abilities over the 6—14 month period.

Keywords: language acquisition; word learning; cognitive development; infancy; psycholinguistics

Introduction

To learn their native language, children must learn words. And to learn words, children must identify words in speech, and grasp what others mean when they talk. The predominant hypothesis about the course of language learning has long been that development proceeds first with speech signal analysis, and only later with discovery of word meaning. This perspective is motivated by demonstrations of precocious phonetic learning between 6 and 10 months (e.g., Jusczyk & Hohne, 1997; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994), subsequent advances in social cognition (e.g., Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998), and finally the onset of referential communication at about 11 months, when infants first produce meaningful speech and gesture (e.g., Bates, Camaioni, & Volterra, 1975; Camaioni, Perucchini, Bellagamba, & Colomnesi, 2004). According to this view, the typical 10-month-old knows the auditory forms of dozens of words, but has yet to invest them with meaning (e.g., Jusczyk, 1997; Jusczyk, 1997; Swingley, 2005), perhaps pending a better understanding of humans as intentional agents.

The notion that development in social cognition is a prerequisite for learning words follows from the premise that the typical conditions under which infants encounter
words are insufficient for infants’ making the connection between the words and their denotations using perceptual association mechanisms alone (Gleitman & Gleitman, 1992). If a parent says “I’ll go get a spoon” in the absence of a spoon, this “teaching trial” is misleading for the simple associative learner who perceives “spoon” and some spoonless applesauce, but is potentially helpful to the intention-reading child who tracks the parent’s goals until he returns with the spoon. Not all researchers agree about this premise, however, maintaining that whatever social cognitive skills infants may or may not have, words and their referents co-occur with sufficient reliability to be learnable by infants using domain-general cognitive capacities for perceptual association. Thus, there is debate about whether intention-reading skills are necessary for young children’s learning of all words (Tomasello, 2001; Waxman & Gelman, 2009), perhaps just “hard”, more abstract words (Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005), or no early words at all (Colunga & Smith, 2005).

One empirical approach to characterizing the mechanisms of early word learning is to test lexical knowledge in children who have only very rudimentary social cognitive skills. Indeed, young infants’ early intention-reading and joint attention skills are limited. For example, at 6–7 months, infants can follow a person’s gaze to an object, but do not appear to understand that gaze implies object-directed interests or goals (Woodward, 2003). Such infants do not yet engage in true “triadic” interactions where they knowingly share attention to an object with another person (Carpenter & Liebal, 2011); the ability to appreciate gaze as both social and goal-directed does not appear until around 9-10 months (Beier & Spelke, 2012). On the other hand, more basic goal attribution and belief computation has been shown in social cognition research around 6–7 months (Csibra, 2008; Kovacs, Teglas, & Endress, 2010).

Despite 6–7-month-olds’ apparent lack of sophistication in recognizing others’ intentions, two recent studies have shown that 6-7 month olds know some object word meanings, including words referring to body parts, e.g. hand, and foods, e.g. banana (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012). Word understanding at this age implies either that rich intention-reading skills are not necessary for learning all words, or that such skills have been underestimated in 6-month-olds.10

Given the theoretical possibility that these object words may have been learned by infants using generic mechanisms of perceptual association and categorization, here we examined more abstract words, such as “eat,” “wet,” and “hi,” whose referents in the child’s experience are, visually speaking, more diverse from instance to instance. Moreover, while concrete words are often used in the presence of the objects they refer to (Gogate, et al., 2000), abstract words such as action verbs have denotations that are often transient by nature, and instances of such words may not be as closely linked in time to

---

10 A third possible response is to stipulate that the 6-month-olds have learned associations, not words, because such infants do not understand reference. We make the contrary stipulation that knowing what a word means begins when infants connect the sound form to a significant, representative aspect of its denotation.
their referents (Tomasello & Kruger, 1992). Learning such words may thus be more challenging for younger infants.

We tested children ranging from 6 to 16 months. This served three goals. First, if 6-9 month olds fail with abstract words, it may indicate that learning concrete and abstract words requires different skills with different developmental courses, or that the learning conditions for abstract words are less favorable. In contrast, if infants succeed, it would suggest that even perceptually diverse categories can be learned and linked to words by children without fully-developed intention-reading skills. Second, if children start to succeed between 10 and 12 months, it will suggest that learning abstract words, unlike concrete words, emerges in parallel with important advances in social cognition (though it would not show that this developmental link is a causal one). Third, if word-understanding performance improves significantly at around 14 months, as Bergelson and Swingley (2012) found for more concrete words, it will provide further evidence for a change in language-relevant cognitive or social abilities in children, including perhaps a deeper understanding of joint attention (Carpenter & Call, in press), a better grip on the conventional nature of words (Buresh & Woodward, 2007), or improvement in appreciating the nature of the experimental task.

To better interpret developmental features of our word-understanding results, we also conducted a series of video-corpus analyses of the contexts in which parents use the concrete and abstract words tested in our studies. Coders annotated a range of interactional features in all instances of these words in 20 recording sessions from the Providence corpus (Demuth, Culbertson, & Alter, 2006). These data were supplemented with analyses of word frequencies in the Brent and Siskind (2001) corpus.

Material and Methods

Participants.

Three age groups were tested: 34 6-9 month-olds (M=8.37mo., R=6.24-9.79mo., 19 girls); 46 10-13 month-olds (M=11.96mo., R=10.02-13.99mo., 26 girls); 18 14-16 month-olds (M=14.99mo., R=14.03-16.52mo., 11 girls). 48 infants were excluded due to fussiness (39), technical problems (3), failure to meet language or health criteria (2), or parental influence (4). Infants were recruited from the Philadelphia area by mail, email, phone, and in person. All were healthy, carried full-term, heard >75% English at home, and had no history of chronic ear infections.

Materials.

Infants were presented with 14 5s videos organized into 7 yoked pairs: all-gone–hi, eat–hug, dance–kiss, more–splash, drink–smile, bye–uh-oh, and sleeping–wet. Videos (each 12x16 cm) were displayed side-by-side on a 34.7x26.0-cm LCD screen (see Fig. 1 and supplementary videos available online).

Items were selected as the most common picturable words, excluding object labels, in young children’s environment based on frequency in a corpus of 16 mothers speaking to their infants (Brent & Siskind, 2001), and based on a database of parental reports indicating which words parents believed their children understood or said (MCDI, Fenson, et al., 1994). Each word appeared in 37-100% of the 16 Brent mothers’ speech,
and had a corpus $\log_{10}$ frequency ranging from 1.5-3.1. Each word was reportedly understood by 38-99% of 16 month olds in the MCDI database.

Fig. 1. Sample stills from video stimuli. The left still is from a trial with videos depicting ‘hi’ and ‘all gone’; the right still is from a trial with videos depicting ‘eat’ and ‘hug’ (left to right, respectively, for each). Video stimuli were in color.

**Apparatus and procedure.**

Visual fixation data were collected using an Eyelink CL computer (SR Research), with a reported accuracy of .5°, sampling monocularly at 500Hz. The eyetracker operated using a camera just below the computer screen, and required no head-restraint. A sticker with a high-contrast pattern, which aided the eyetracking mechanism, was placed on the infant’s forehead.

Before the experiment began, the procedure was explained to parents, who gave informed consent. Parents completed a vocabulary checklist and a word exposure survey estimating how often their child hears our test words in daily life. Then, parent and child were led to the dimly-lit testing room where the infant sat on the parent’s lap facing a computer display. Parents wore an opaque visor preventing them from seeing the screen, and headphones over which they were prompted with the target sentence.

Each of the 7 yoked pairs of videos was presented four times, resulting in 28 test trials. On each test trial, parents spoke a single sentence to their child, repeating a prerecorded utterance that they heard over headphones. The sentences had been recorded by a native English-speaking woman talking at a moderate speed, with slightly exaggerated intonation. Each sentence followed the format “Look! X, X!” where X stands for the target word. The recorded sentences were 3.5s in duration and were presented at about 34dB, audible only to the parent. The exact timing of parental sentences varied across trials, but the onset of the target word was recorded; the videos played for ~3s after the parent’s utterance ended.

Each test trial began with a beeping, spinning star that drew children’s attention to the screen’s center. Once the child fixated it (or after 10s), the pair of test videos was shown twice, accompanied only by music. This familiarized children to the videos and their locations. Then the videos were shown again, twice, with the mother being prompted to name one of the videos during the first of these two presentations.
Each participant was randomly assigned to one of two pseudorandomized trial orders, with target side counterbalanced. All children were tested on all 14 items. The experiment lasted about 20 minutes. Families were compensated with a choice of two children’s books or $20. The entire visit lasted about 45 minutes.

**Corpus Analyses**

We examined mothers’ use of the words tested here as well as the words tested in Bergelson and Swingley (2012) in both the Brent Corpus (an audio corpus of 16 mothers interacting with their 9-15 month old infants), and in 20 videos of the Providence Video Corpus (5 mothers interacting with their young children; we selected a subset in which children ranged from 11 to 18 months). In the Brent Corpus we compared frequency counts in isolation (i.e., in one-word utterances) and overall. In the Providence corpus we extracted 919 utterances in which both the mother and child were clearly visible, and in which one of our words of interest was said. These utterances were coded for a number of features, including whether the referent of the word was present (e.g. is there an apple when ‘apple’ is said, is someone eating when ‘eat’ is said, etc.), what the parent was looking at/touching, what the child was looking at/touching, the situation the word was used in, what (if anything) was moving, whether the word was said before, during, or after attention to the relevant referent transpired, and what was present in the room. In the case of body-parts, which were evidently always “present” during every interaction, coders noted “presence” only when the relevant part was, in any important sense, involved in the interaction: for example, if the mother was feeding a child who had yogurt all over her mouth and said, gazing at her, “look at your messy face!” this counted as “presence” of the word “face”; in contrast, if the child was crying and the mother was holding and hugging him while singing “if you’re happy and you know it clap your hands”, this did not count as an instance in which “hands” were considered “present”.

**Results**

**Results from eyetracking study.**

To measure whether infants fixated the named event more upon hearing it named, we computed a difference in fixation proportions: how much infants looked at one video when it was the target, minus their proportion of looking to it when it was the distracter. This computation, which corrects for bias due to preferences for one video over the other (Bergelson & Swingley, 2012), yields one score for each item-pair. For instance, with the pair *kiss–dance* an infant’s performance was given as how much he looked at ‘kiss’ when it was said by his parent, relative to his looking at ‘kiss’ when ‘dance’ was said. Positive difference scores indicate word understanding.

We measured performance in the window from 367-4000ms after the onset of the spoken target word (e.g. the beginning of the first ‘hi’ in “Look! Hi, hi.”). Fixation responses earlier than 367ms are unlikely to be responses to the speech signal (Swingley, 2009). The 4000ms window offset is used here, rather than the 2000ms offset typically used with children over 18 months, because younger children take longer to demonstrate word recognition (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998).
Fig. 2. Target looking performance in each infant. Data are subject mean difference scores calculated over the 367-4000ms window. These were calculated by averaging the 7 item-pair mean difference scores for each subject. Symbols indicate the age bins used for statistical analyses; see text for details.

Analyses of children’s fixations revealed no indication that 6–9 month olds understood the words we tested. 19/34 infants showed a positive proportion of target looking (see Fig. 2; Mdn=.020, p=.47 all Wilcoxon tests unless noted otherwise). Performance on 4/7 item-pairs was positive (see Fig. 3; Mdn=.027, p=.77). By contrast, 10–13-month-olds looked at named targets significantly above chance levels, both over subjects and item-pairs. 32 out of 46 infants showed a positive proportion of target looking (Mdn=.075, p=.002, binomial p=.011). These infants showed positive performance on 6/7 item pairs (Mdn=.060, p=.02). Finally, 14–16 month olds showed consistently high levels of performance. 15/18 infants showed positive increases in target looking (Mdn=.12, p=.0017; by binomial test p=.0075.), with positive performance on all 7 item-pairs (Mdn=.14, p=.0078).

A correlational analysis found no relation between children’s performance and the total number of words parents reported that children understood or said, except in the
eldest group, as determined by the MacArthur-Bates Communicative Development Inventory (CDI; $\tau=.43$, $p=.012$; all other $p>.1$ by Spearman (nonparametric) correlation test). Considering parental report of children’s knowledge of the specific words tested in the study, again only in the eldest age group was this vocabulary knowledge correlated with gaze performance ($\tau=.17$, $p=.016$; all other $p>.1$).

Similarly, an analysis examining parents’ estimates of the frequency with which a child hears the study’s words in his or her daily life (on a 0-4 scale ranging from ‘never’ to ‘several times a day’) found no relationship between this measure and children’s word-recognition, except in the eldest group ($\tau=.20$, $p=.006$). Descriptively, most parents in all age groups said their children heard all of our test words ‘several times a day’; this response was more frequent than all of the others combined ($M=3.3$, $SD=1.06$).

**Concrete and abstract word comparisons.**

As a further comparison with previous research on concrete nouns (Bergelson and Swingley 2012), we statistically compared subject means in that work and in the current experiment, for each age group. 6-9 month olds did significantly better on concrete words than on abstract words, with 26/33 infants achieving positive subject means, compared to 18/34 here (estimated difference=.064, $p=.012$ by Wilcoxon test, Chi Square=3.88, $p=.049$). 10-13 month olds did not show significantly different performance on the two word types, with 20/30 infants achieving positive subject means, compared to 34/46 here (estimated difference=.0096, $p=.83$ by Wilcoxon test, Chi Square=.18, $p=.67$). 14-16 month olds showed marginally different performance, with 7/7 attaining positive subject means, compared with 15/18 here (estimated difference=.17, $p=.055$, Chi Square=.22, $p=.64$).

Additionally, a series of analyses was conducted to test whether the difference in performance between abstract words (shown here) and more concrete words (Bergelson & Swingley, 2012) might be due to higher frequency of the concrete words rather than something more fundamental about the words’ meanings. Frequency was estimated using the Brent corpus (Brent & Siskind, 2001). There was not a significant difference in the frequency of the abstract and concrete words. Descriptively, concrete words occurred 45-562 ($M=262$, $Mdn=244$) times within the corpus while abstract words occurred 33-1292 ($M=453$, $Mdn=219$) times. Across each set of words, the total number of usages did not vary significantly (244 versus 219, $p=.98$ by Wilcoxon test). Given that previous research supports a link between word learning and frequency of isolated word tokens (Brent & Siskind, 2001), we also examined this variable here. The sets of words were not differentially likely to occur in isolation (in single-word utterances) either: concrete words occurred 2-92 ($M=26$) times and abstract words occurred 0-1091 times ($M=152$); this difference was not significant (concrete $Mdn=19$, abstract $Mdn=11$; $p=.95$ by Wilcoxon test.).

Analyses of the Providence Corpus (Demuth, et al., 2006) revealed that there too, our abstract and concrete words occurred with similar frequency: abstract words occurred 1-94 times ($M=37$, $Mdn=23$), there were 523 abstract-word tokens total. Concrete words
occurred 5-46 times (M=21, Mdn=19), with 396 concrete-word tokens total (estimated difference per word type: 7 words; p=.29 by Wilcoxon test over words). Similarly, abstract and concrete words as a group did not differ in number of isolated occurrences (72 isolated abstract-word tokens total, R=1-7 over words; 35 isolated concrete-word tokens total, R=1-3 over words; estimated difference 1.8 words; p=.13 by Wilcoxon test over words).

Hand-coding of interactional features during parental use of the tested words revealed a large word-type (object versus action) difference in whether the referent of the word was present as part of the interaction. Abstract words were said much more often than concrete words when their referent was not present—e.g., saying “hi!” when no-one was newly on the scene, or “kiss” when there were no evident attempts at kissing. By contrast, concrete words (“a banana!”) were more often spoken in the presence of the referent (an actual banana, or a picture of one). For abstract words the referent was not present 39% of the time; for concrete words, 15%. This pattern held for 5/5 children in the corpus, and was significant over words (estimated difference =.24, p<.012 by Wilcoxon test over words).

No significant differences between abstract and concrete words was found in what mothers or children were touching or looking at, the number of situation-types that the word occurred in (e.g. playing, eating, interacting, book-reading), what in the scene was moving (e.g. child or mom, their hands, other objects, etc), whether the word was said before, during, or after attention to the relevant referent transpired, nor what was present in the room (all ps>.05 by Wilcoxon tests, and not significant predictors in logistic regressions of word-type). In short, on most coded variables, abstract and concrete words did not differ in various features of the learning environment.

Discussion

These findings enrich our understanding of the early stages of language acquisition, showing that by 10-13 months, but not earlier, infants linked several common abstract words to their referents. This in turn suggests that the word-learning mechanisms and social/cognitive abilities that are needed to learn abstract words under ordinary daily-life conditions are in place approximately half a year earlier than previous laboratory tests had indicated (Golinkoff, et al., 1987). At the same time, the results are consistent with diary and other observational studies of children of 10 months and older. Such studies have found that a wide range of word types is present in children’s early comprehension vocabularies (e.g., Bloom, 1993; Dewey, 1894; Nelson, 1973). For example, Benedict (1979) found that when infants appeared to know 10 words, at around 10 months, among those words were nominals, action words, and various social words. The present research substantiates these claims using a controlled, replicable experimental procedure. Recognition of words in our study is particularly remarkable because it required that infants generalize the words they knew from their individual life experience to new instantiations involving actors and events previously unseen by the infant.

The failure of the 6–9 month olds to evince recognition of non-object labels contrasts with performance of 6–9 month olds in other studies testing understanding of
object words. Those studies used similar fixation-based methods (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012) or event-related potentials (Parise & Csibra, 2012), so it is unlikely that minor methodological differences between this study and prior studies account for the difference. Nor is lexical frequency likely to be responsible, given results of the corpus analyses described above.

A more likely possibility is that the developmental difference concerns the requirements for learning more abstract words, for which the connection between the uses of the words in conversation and the concepts to which they refer is more difficult to establish through observation (Gillette, Gleitman, Gleitman, & Lederer, 1999). This hypothesis has two versions. One is that the same learning machinery is at work in learning concrete and abstract words, but the statistics of abstract words are more complex and therefore demand more data to resolve, which is manifested here in the later age at which evidence of learning is found. The other is that learning abstract words demands skills that do not begin to emerge until around 10 months, such as the capacity for reading others’ intentions. For example, between nine and ten months, infants improve in their ability to analyze the gaze of others as social, goal-directed action (Beier & Spelke, 2012; Brooks & Meltzoff, 2005). This ability could prove more useful for understanding abstract words than concrete words, because abstract words may have fewer correlated perceptual features. That is, shape, size, color, movement, and texture range less freely for things called ‘juice’ than for situations called ‘all gone!’ or ‘uh-oh.’ While our task does not itself require gaze-following, abstract words are, to a greater degree, expressions of the parent’s perspective, and as such their learning might depend more on skills of intention-reading.

Moreover, concrete words and referents are more easily subject to joint visual attention in a way that abstract words may not be: objects are often present in the environment before, during, and after the words labeling them are uttered, and thus reading others’ visual attention is helpful when linking object words to referents. In contrast, many common verbs toddlers hear are transient. Indeed, previous research has found that common verbs are usually said before the relevant action occurs. Because these verbs do not co-occur with the action they denote, it has been suggested that to learn their meanings “children must find cues other than ostensive gestures to determine the adult’s attentional focus” (p.313, Tomasello & Kruger, 1992).

On the other hand, the data-driven account of development may gain support in our finding that abstract words are less likely to be said when the referent is evident in the context of the interaction. In principle, this could make learning abstract words harder even if they are learned through the same mechanism as concrete words.

A related explanation for the delay, in keeping with the data-driven hypothesis, is that infants first attempt to interpret words as names for concrete objects, and only upon failing this do they attempt to link them to actions, or other more abstract concepts, a misstep that lengthens the process of learning abstract words. Of course, this presupposes the existence of an object bias, and at present there is no evidence that speaks to this issue in 6-9 month olds.

Our results do not show which of these hypotheses is correct, but they do indicate a developmental change that requires explanation.
Infants’ performance in the task improved substantially at around 14 months, just as in Bergelson and Swingley (2012). This change, which is evidently independent of whether the tested words are abstract or concrete, might be due to increases in basic cognitive capacities. As discussed in Carpenter and Call (in press), a mature form of joint attention in which an infant tracks not just what she knows, or an adult knows, but what they both know together, emerges around 14 months; joint attention demonstrations prior to this age may lack this “knowing together” element (p. 7, Carpenter & Call, in press). Additionally, infants around this age, but not around 9 months, are able to track both (a) that an individual reaching for something likes it, though this preference does not necessarily apply to others, and (b) that if an individual uses a label for an object, another individual is likely to mean that object by that label as well (Buresh & Woodward, 2007). Both mature joint attention and insight about word-label generalization may be among the skills that improve around 14 months and that lead infants to perform better in word comprehension tasks with both concrete and abstract words.

Parental reports of infants’ vocabulary knowledge did not correlate with infants’ gaze performance, just as in Bergelson & Swingley (2012). This might be attributable to the difficulty parents have in assessing their infants’ vocabulary knowledge before children begin to talk and while children’s responses to language may be quite ambiguous. This assessment on the part of parents may be particularly difficult when considering words that do not refer to objects.

Conclusions

The current findings contribute to the literature on language acquisition in several ways. We showed that infants as young as 10 months old identify novel referents of common words that do not refer to concrete objects, but younger infants do not. Thus, the acquisition of abstract and concrete words differs ontogenetically, and may require skills with differing developmental trajectories. The beginnings of abstract word learning, but not concrete word learning, appear to occur in parallel with the major advances in social cognition documented in prior research (e.g., Carpenter, Nagell, Tomasello, et al., 1998), though word-referent consistency likely plays a role in the more protracted timeline of abstract word learning as well. Furthermore, we replicated with abstract words what has been shown with more concrete words: that at around 14 months infants’ word-learning or word-recognition abilities improve greatly. This improvement too coincides with important improvements in social cognition found in the literature (e.g., Buresh & Woodward, 2007). Taken together, these findings show that infants’ early word learning comprises various types of words; involves generalization over prior experience in non-obvious ways; and is characterized by two developmental shifts: one around 10 months, and one around 14 months.

Acknowledgements This work was funded by an NSF Graduate Research Fellowship to EB, the NSF-IGERT Program, and NIH Grant R01-HD049681 to DS.
Chapter 4: Young Infants’ Generalizations over Word-Forms during Comprehension

Introduction

One critical aspect of word comprehension is recognizing words across a range of contexts, in which subtle or gross acoustic differences alter how words sound. Successful, proficient, and efficient word comprehension requires a good understanding of a given language’s phonology, such that words are recognized in all and only intended instances. This in turn requires a stored lexical entry for a given word that can be referenced against incoming acoustic input. However, it is not entirely clear how such representations are built in the growing lexicons of infants; this is the question we turn to here. That is, in this chapter we ask whether infants’ early word comprehension reflects word-form representations that are subject to the same sorts of generalizations as those of adults.

Closely tied to this question is the timeline of acquisition of native language phonology. In a typically developing infant, language-specific phonological development can be found around 6 months, at least for a certain (but perhaps limited) range of vowel contrasts (Kuhl, et al., 1992; Polka & Werker, 1994). Before that, infants have a language generic pre-phonology: they rely on universal discrimination abilities, and must learn which kinds of acoustic differences matter, linguistically, and which do not. This universal discrimination begins to narrow, with more language-specific sensitivity emerging by around the first birthday (Kuhl, et al., 2006; Werker & Tees, 1984, inter alia).

Until recently, it was thought that a still-solidifying phonology did not pose a problem for word-learning, since word-learning was not thought to take place until about 12 months, when infants’ language-specific phonology is more stable (e.g. Kuhl, 2011). In contrast to this assumption, in recent work we have shown that as early as 6 months, infants understand the meaning of many words for foods and body-parts, when said by their parent (Bergelson & Swingley, 2012, hereafter B&S12). Others too have found that infants understand proper nouns and the words ‘hand’ and ‘feet’ at 6 months (Tincoff & Jusczyk, 1999, 2012). This recent literature, in turn, allows us to next raise questions about the nature of these early pre-phonological word-form representations.

Given that infants’ speech perception is still in flux at the age that word comprehension is beginning, it is possible that infants at the earliest stages of word comprehension have word-form representations that differ from those of adults. Indeed, infants must discover various generalizations about what ‘counts’ as an instance of a word. For instance, infants must come to understand that the relatively large acoustic differences generated by indexical properties, i.e. changes in voice quality, affect, speech rate, etc., do not change the identity of words. It is as yet unknown whether, early on in word learning, infants have realized this, or, whether they instead limit the word-forms they recognize to those that are most similar acoustically to the word-forms they hear most. For example, at the extreme, we could imagine that an infant only recognizes the
word ‘dog’ if it is spoken by her babysitter, and not by anyone else, if that is the only person she has heard say that word. On the other hand, it is possible that early on, infants interpret the speech they hear into some sort of normalized phonetic space that allows them to recognize word tokens from new speakers with ease. In the intermediary case, one could imagine that word comprehension is worse with a new talker, but still successful. In Experiment 1 we investigate these issues by asking whether word comprehension is impaired when infants are exposed to a new talker. Importantly, we are not asking whether infants notice a change of speaker or affect, or generalize across tokens of a given unknown word that varies across these indexical changes, but rather, we are looking at comprehension in the context of a change in talker.

Another way that infants have to make the right generalizations about word-forms is that they must realize that phonemic changes, even subtle ones, may change a word’s meaning (cf. ‘smile’ and ‘smell’). Here too it is possible that at the outset of word comprehension infants do not know this. That is, they may have a fuzzy notion of the sound structure of a word, and thus accept words that sound ‘close enough’ as instances of a word, e.g. accepting ‘beel’ for ‘ball’. On the other hand, infants may show a high degree of phonetic specificity in what they consider to be instances of words from the onset of word learning onward, either because they have a high degree of precision in their lexical specification for a word, or because, as discussed above in the indexical variation case, they are constrained in the acoustic variation they will accept as a ‘match’ from previously encountered tokens. Experiment 2 takes a first step in looking at word-form precision by asking whether word comprehension is impaired if the word being spoken is altered by a single phoneme, namely, its stressed vowel. Here too, we look at mispronunciation in the context of word comprehension, rather than determining whether infants can discriminate e.g. ‘bee’ from ‘boo’.

Thus, in the two studies presented below we query two types of word-form generalizations in infants early word comprehension: we test infants’ speaker invariance, (Experiment 1) and word-form precision (Experiment 2). Note that for infants to show adult-like competence in these experiments, they must in one case find that an acoustic difference does not matter for word meaning (i.e., in Experiment 1, where the uttered word should still trigger the relevant lexical item despite the change in talker), while in the other case they should find that an acoustic difference does matter for word meaning (i.e. in Experiment 2, where a phoneme change should negatively impact word comprehension).

In both studies, we looked at a range of ages, beginning at 6 months, to determine whether and how these abilities improve over time. Previous work in each of these domains, reviewed below, has by-and-large looked either at older infants only (e.g. 12-20 months), or if it has looked at 6-11 month olds, has used paradigms that query infants’ discrimination and generalization abilities between two speech samples, without necessitating that infants understand the words in question (for an exception, see Parise & Csibra, 2012). Given that infants’ speech-sound discrimination abilities do not always reflect whether they can map words onto objects (Stager & Werker, 1997), it is important to query word recognition abilities in a task that looks not just at infants’ speech processing, but at the link between the way a word is said and its meaning. This pair of
studies does just that by examining whether infants appropriately constrain their word-form representation during spoken word comprehension. By using the same materials as B&S12, these studies provide a between-subjects comparison of three cases: infants understanding of words when spoken by their mother (B&S12), spoken by an unfamiliar person (Experiment 1), or mispronounced by their mother (Experiment 2). The comparison across these studies will provide some first steps in our understanding of how speaker- and phonetic-specificity interact at the beginning of word comprehension.

**Experiment 1: Speech Invariance**

In this experiment we assessed first whether infants understand words for foods and body parts when the words for these items are said by an unfamiliar person, and secondly, whether this understanding was diminished or delayed with respect to infants’ understanding when the words were said by their mother (B&S12). Thus, this experiment investigates whether early word representations are perhaps more exemplar-based, that is, more tied to the precise acoustic word-forms infants hear most often from their parents, or whether they are robust to the kinds of individual differences we find across speakers from the outset of word comprehension onward. That is, we examine whether infants exhibit speaker invariance during early word comprehension.

The kind of speech invariance investigated in Experiment 1 is not context-induced variation in different phonetic environments, but rather indexical variation in speech, due to individual talker differences (Perkell & Klatt, 1986). These differences are caused both by physiology, e.g. vocal tract length, and by mood, speech-rate, attention, and other variables that, in the general case, do not change the meanings of words. As competent language users, we have the ability to both encode these superficial differences, noting for instance, if someone has a deep voice, or to hear through these sorts of differences, in understanding what people say despite differences in voice quality or speech rate, within reasonable limits.

This everyday ability is actually quite a complicated one, and the mechanisms that underlie it are poorly understood, both from a psychological perspective, and from an engineering perspective. For instance, while speech recognition systems have gone through monumental improvements in the past decades, even top-of-the-line systems still have high error rates beyond the single word domain, and generally rely on training from a specific speaker’s input over time to provide correct responses. One estimate finds Apple’s Siri and a Windows competitor have a 12-20% error rate at the one-word level, but a 70-85% error rate at the sentence level (Fradrich & Anastasiou, 2012).

Psycholinguistics research with children and adults has shown that although these groups do of course understand words spoken by others very efficiently in daily life, there is nevertheless a cost in spoken word comprehension from multiple speakers as opposed to a single speaker (e.g. Mullenix, Pisoni, & Martin, 1989; Ryalls & Pisoni, 1997). However, in other ‘easy’ processing tasks (such as lexical decision with nonce words that are not very word-like, or shadowing tasks) effects of indexical variation, namely talker identity and speech rate, are not found (McLennan & Luce, 2005). Other work has highlighted that adults’ word-form representations are not completely abstract, and that rather than a detail-free purely phonemic representation, certain aspects of
indexical properties are indeed encoded in memory (e.g. Goldinger, 1997), and referenced during word comprehension (Creel, Aslin, & Tanenhaus, 2008).

Infant research has looked at the speech invariance problem using several paradigms. Using the head-turn preference procedure, the results have been mixed. On one hand, van Heugten and Johnson (2012) recently found that at 7.5 months, infants trained on words embedded in speech by their mother preferred those same words in isolation spoken by their father. Infants also succeeded when the mother and father they listened to in the study were not their own. On the other hand, Houston and Jusczyk (2000) found that 7.5 month olds could not generalize words across male and female speakers, but 10 month olds could.11

Related work by Singh and colleagues found that on a similar timeline, infants exhibited a limited ability to recognize words across affect: at 7.5 months, infants failed to recognize words that were familiarized in one affect (neutral or happy) when they were tested in a passage using the opposite affect, instead showing recognition of the words only in affect-matching passages; by 10 months an affect change does not impair recognition (Singh, et al., 2004). In further work along the same lines, examining variability and affect, Singh (2008), suggests that the specific training and testing contrasts can alter whether infants are more attuned to phonetic or suprasegmental differences in input, and that moreover, exposure to more variable forms during training (or indeed, life experience) leads to more robust word representation.

In a more straightforward examination of whether infants recognize word-forms uttered by others, two studies have looked at early word comprehension using novel speakers. Tincoff and Jusczyk (1999) used a synthesized, gender-ambiguous child’s voice in their study querying infants understanding of the words ‘mommy’ and ‘daddy’ and found that even with this somewhat strange talker, 6 month olds succeeded in recognizing and understanding the words. Using electroencephalography (EEG), Parise and Csibra (2012) attained results that potentially conflict with Tincoff and Jusczyk’s (1999). They found that when a spoken word for an object did not match a visual display, 9 month old infants showed an N400, the neural signature of an unexpected stimulus, but only if the speaker was their mother, not when it was an experimenter. This result suggests that infants’ assessment of a mismatch between an image and a word (which is a proxy for word comprehension) varies as a function of speaker identity.

The inconsistency between these two studies’ results has a few possible sources, some of which are methodological in nature, and some of which are related to word-form representations. On the methodological side, it could be that the studies differed in the recognizability of the new speaker’s voice, familiarity and comfort in the testing situation, the nature of the paradigm, outcome measure, timing of response, etc. On the other hand, this difference could highlight that the words ‘mommy’ and ‘daddy’ are

11 These different patterns of results are explained by van Heugten & Johnson as likely stemming from two factors: first, the speakers in Houston & Jusczyk were voice actors rather than parents of young infants; second, in the older study, target words were presented in list form during familiarization while in van Heugten & Johnson study, they were part of fluent speech in passages. These factors were said to increase the naturalness of the task and allow for success at the earlier age.
stored more robustly, either in terms of the strength of the semantic representations or word-form representations. We address some of these possibilities in Experiment 1 below.

In slightly older infants, using the switch task (Stager & Werker, 1997), Rost and McMurray (2009, 2010) found that increasing the number of speakers labeling novel items led 14 month olds to successfully map words to objects while manipulation of voicing cues within a single speaker did not, even though both kinds of changes are within a natural range of differences for the way a word is instantiated. Using just a single speaker in this paradigm also did not lead to successful mapping (Stager and Werker, 1997). These findings suggest that having multiple talkers utter a word helps with word-learning in a way that phonetic variability, or learning from a single speaker does not. This in turn suggests that increasing indexically-based variability in the tokens that form infants’ word representations leads to better word-form learning. This is similar to Singh’s suggestion that word-form variability helps infants build more robust representations (Singh, 2008), though in Rost & McMurray, the role of variability is assessed in the context of learning a word-referent pairing rather than in a discrimination paradigm.

Thus, the previous literature, while sometimes conflicting, in general suggests that by 10-14 months infants are able to abstract across various aspects of idiosyncratic speaker information when recognizing words or mapping them onto novel objects (Houston & Jusczyk, 2000; Rost & McMurray, 2009, 2010; Singh, et al., 2004). Before this age, the results are a bit mixed, and seem to depend on what aspect of speaker invariance is queried and how (Houston & Jusczyk, 2000; Parise & Csibra, 2012; Singh, et al., 2004; Singh, White, & Morgan, 2008; Tincoff & Jusczyk, 1999; van Heugten & Johnson, 2012). Here we hope to clarify infants’ early knowledge of speaker invariance by examining it in a more naturalistic context: during spoken word recognition of common nouns.

**Methods**

**Participants.**

Subjects were 47 6-12 month olds, split into a younger and older group for the ease of comparison with other studies. The final sample consisted of 15 6-8 month olds (M=7.55mo, R=6.01-8.80mo) and 32 9-12 month olds (M=11.23mo, R=9.2-12.87 mo). Infants were recruited from the Philadelphia area by mail, e-mail, phone, and in person. All children were healthy, carried full-term, and heard 75% or more English in the home. None had a history of chronic ear infections. An additional 42 infants were run, but were excluded from the final sample for the following reasons: technical problems (n=6), language exclusions (n=7), pilot subjects (n=3), general fussiness (n=4), and not contributing data to at least half of items tested (n=21). All parents gave written informed consent on behalf of their infant, in keeping with the IRB procedures at the University of Pennsylvania.

**Materials.**

Materials were identical to those B&S12; see Chapter 2.
Apparatus and procedure.

The basic experimental setup and structure in this study was identical to that in B&S12, with one crucial change. Instead of the mother wearing the headphones over which she heard the sentence prompt which she repeated to her child, now an experimenter wore the headphones and repeated the sentence aloud to the child, from behind the experimenter’s computer; the infant remained in her mother’s lap, and the mother, as in B&S12, wore an opaque visor through which she could see her child but not the display presenting the stimuli. The experimenter too could not see the display presenting the stimuli to the infant. See Figure 1.

Figure 1: Experimental Setup. In this Study, infants sat in their parent’s lap, and experimenter sat behind a second computer. The experimenter heard sentence prompts over headphones that she then repeated to the child. Neither parent nor experimenter could see the images displayed to the infant. The setup in B&S12 was identical, except that the parent wore the headphones rather than the experimenter.

Results

We analyze separately the results from each age group (6-8 and 9-12 month olds) and trial-type (paired picture trials & scene trials), using the same analyses, data processing and outcome measured described in greater detail in B&S12. Briefly, paired-picture trials showed two images on a grey background, one food, and one body-part, while scene trials showed either a whole person or face, or four food-related items on a tabletop. Paired-picture trials were subject to a difference score analysis, which looked at the proportion of time (over the 367-3500ms window of interest) that infants looked at an image when it was the target as opposed to when it was the distracter, resulting in item-pair level data. For the scene trials, the outcome measure was the proportion of target looking, corrected for baseline looking, over the 367-3500ms window of interest. Trials only entered these analyses if, in a given trial, infants looked at more than one interest area, and looked for more than 1/3 of the window of interest. Infants who did not contribute data to at least half of the pairs (paired-picture trials) or items (grouped trials)
were excluded from analyses. Note that these criteria for subject exclusion were
retroactively applied to B&S12, so that the samples compared below were the result of
the same selection process.

**Experiment 1, paired picture trials.**

Infants in both age groups showed strong performance on the paired-picture
trials; see Figure 2, middle panel. 12/15 6-8 month olds looked longer at the correct
image, resulting in a significantly positive difference score across subjects\(^{12}\) (M=0.033,
p=0.047; p=0.035 by binomial test). Over item-pairs, 6-8 month olds showed positive
difference scores on 6/8 item-pairs, though performance over item-pairs did not reach
significance (M=0.032, p=0.16).

23/32 9-12 month olds showed positive difference scores (M=0.055, p=0.017;
p=0.021 by binomial test). Over item-pairs, 9-12 month olds showed positive difference
scores on 7/8 item-pairs, which reflected performance significantly above chance over
items (M =0.066, p=0.027). Taken together, these results demonstrate that infants
understood food and body-part words in the paired-picture trials, both at 6-8 and at 9-12
months of age.

For paired-picture trials, 6-8 and 9-12 month olds’ performance did not differ
from each other by two-sample Wilcoxon Test (estimated difference=0.0018, p=.94). This
was also the pattern found in B&S12 across the 6-13 month age range.

**Experiment 1, scene trials.**

Infants’ performance on scene trials was at chance levels, both over subjects, and
over items, both in the younger and older group of infants. 3/12 6-8 month olds and 11/32
9-12 month olds showed positive corrected proportions of target looking (6-8: m=0.029,
p=.87, 9-12: m=0.0042, p=.66).\(^{13}\) Over items, 6-8 month olds showed positive
performance on 6/16 items, 9-12 month olds on 8/16 items (6-8: m=0.0054, p=.67, 9-12:
m=0.0065, p=.55). The performance over subjects across the two age groups did not
differ (estimated difference=0.015, p=.51).

**Comparison with B&S12.**

Showing that infants succeed when hearing the experimenter label the images still
leaves open the possibility that they nevertheless showed worse performance than when
their mother said the words. We thus compared performance in the original data set with
the present results (see Figure 2).

---

\(^{12}\) Given that the data are not normally distributed, unless stated otherwise, the statistical test used is a
Wilcoxon Test. For group comparisons against chance, one-sample one-tailed tests are used, given that
only positive results are expected. For cross-group comparisons, two-tailed two-sample tests are used
across groups of subjects; paired two tailed tests are used across items and item-pairs.

\(^{13}\) Please note that there are 3 fewer participants in the younger age group for the scene trials because in
these trials more infants were excluded for providing data on less than half of items based on our criteria
that they look for one third of the window of interest, and at more than one interest area.
For both trial-types (paired-picture and scene trials), in both the younger and older group, over item-pairs and over subjects, performance did not differ significantly from an age-matched subset of B&S12 (for all pairwise two-sample comparisons, p>.17). We return to an in-depth analysis and discussion of these comparisons in the results section following Experiment 2, where we compare all three studies.

<table>
<thead>
<tr>
<th>B&amp;S12: Mother Talking</th>
<th>Exp. 1: Talker Change</th>
<th>Exp. 2: Vowel Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>ns</td>
<td>*</td>
</tr>
<tr>
<td>17/22</td>
<td>12/15</td>
<td>12/17</td>
</tr>
</tbody>
</table>

Figure 2: Subject means for B&S12 and Experiment 1 and 2. Each panel shows each subject’s mean difference score across item pairs for paired-picture trials (see text and B&S12 for details). The three panels, from left to right, show the data from the original B&S12 study, Experiment 1, and Experiment 2. Stars with brackets indicate significant performance over the bracketed age range, stars under straight bars indicate significant performance across bracketed groups, see text for further details. Ns stands for ‘not significant’, and the fractions indicate the number of subjects with positive means for each panel in the 6-8 month age range.

**Experiment 1 Discussion**

**Paired-picture trials.**

The results from the paired-picture trials in Experiment 1 suggest that at the same age that infants demonstrate knowledge of words for foods and body-parts, i.e. at 6 months, infants are able to appropriately generalize across voices, recognizing these words to similar degrees when they are said by their parent or by a new individual. These results suggest that infants show appropriately broad links in what they take to be instances of words they know, even when these tokens vary from those they likely hear most often from their parents.

One potential limitation of this study is that experimenters’ voices were too similar to mothers’ to really test the limits of infants’ speech invariance abilities. Both the experimenters’ and mothers’ voices varied to a great degree, both within and across subjects, though the experiment was not designed to quantify these differences. However,
one natural feature of our data allows us to gain traction on how robust infants’ speaker invariance is. Approximately one third of the younger group of infants in this study were African American, and in keeping with the demographics of Philadelphia, were by-and-large low-income learners of African American Vernacular English who lived in West Philadelphia, a predominantly African American neighborhood. Thus, for these infants, more than for the others, the experimenter’s voice may have varied maximally from their mothers’, given that our experimenters were all Caucasian or Asian, middle- to upper-middle class speakers of the regional white dialect.

When we examine the African American subset of the data, though the number of subjects is small, performance was still significantly above chance even in the 6-8 month old group; in fact, all subjects showed positive difference scores (n=5, paired-picture trials, m=.079, p=.031). Performance did not differ significantly from the African American Subset of 6-8 month olds in B&S12 (n=14, estimated difference=-.001, p=.96). 14

While this analysis confirms that infants succeeded even with voices that differed quite a bit from their mothers’, it is nevertheless the case that the kind of abstraction over indexical variables required in this study is somewhat modest. Indeed all of the experimenters’ tokens likely sounded, by any acoustic or phonetic metric, more like infants’ mothers’ tokens of the target words than like the words for the other concurrently displayed images. At the same time, the kind of talker differences probed in this study are certainly within the range of speech invariance infants must cope with on a daily basis. Thus, this work provides an important first step in understanding infants’ word comprehension in the face of indexical changes. Here we have replicated our original findings of early comprehension with a new group of 6 month olds, and extended those results to include comprehension of words spoken by a new person. It is up to future work to take these findings further, looking at other indexical variables such as speech rate, affect, gender, age, etc. For instance, recent work suggests that recognizing word-forms across new voices may be easier than recognizing word-forms from accented speakers (Schmale & Seidl, 2009), and moreover, that understanding accented speech may be challenging until well into infants second year (Best, Tyler, Gooding, Orlando, & Quann, 2009). Given the results of Experiment 1 and B&S12, further work could begin probe the limits of speech invariance in the first year of life. At this point, what we can say is that the differences between speakers did not seem to impede understanding in early word learners. This in turn suggests that infants’ early word comprehension for food and body part words in this task is robust to superficial acoustic differences that are not contrastive for meaning.

Scene trials.

While infants’ performance on scene trials did not indicate word knowledge, it also did not differ from their performance in B&S12, where their mother uttered the words. There too, performance in these trials types was fragile: while the majority of

---

14 This pattern holds for the older group of infants as well.
infants succeeded over a wider age range than we examine here, performance was especially robust only after 13 months, which is beyond the age range used in the present study. We feel the data here are weaker for several linguistically uninteresting reasons, related to the difficulty inherent in these more complex trials, complications in how these visually complex trials must be analyzed, and saliency issues in the scene trial materials. We discuss these issues further in Appendix A. Given that performance in the scene trials, just as in paired-picture trials, did not differ based on whether the speaker was the infant’s mother or the experimenter, we believe that these results, though negative, provide no evidence that infants’ understanding of words for foods and body parts is diminished when these words are uttered by a new speaker.

**Underlying mechanism.**

Our findings, like those of Tincoff & Jusczyk (1999) show that by 6 months infants are able to understand words said in a voice that is not their parents’. In our case, we can directly compare these results to B&S12 and say that they understand words just as well as when they are uttered by infants’ own mother. However, these results seem to conflict with those by Parise & Csibra (2012), who find that 9 month olds do not show an N400 (the neural marker of hearing or seeing something unexpected by the context) when there is a mismatch between the word they hear and the image they see, unless the speaker is their mother. It is not entirely clear why our results would differ, but given that the nature of ERP and eye-movement responses is quite different, it is possible that these methods are tapping different aspects of word-form processing.

For instance, the N400 may be more responsive to tokens that are more similar to stored tokens, such that a response to a new speaker’s token may be more diffuse or noisy, not rendering a clear N400. Another possible explanation is that the specific utterance timing across the studies differed, in that the experimenters’ utterances in Parise and Csibra were yoked to other mothers’ utterances in the study, and thus their timing may have been detectably unaligned with the infant’s attentiveness. A third alternative is that the words chosen by Parise and Csibra have more tenuous word-form representations for 9 month olds than those used in the present study or by Tincoff and Jusczyk (1999), and thus infants are better at recognizing their mother’s utterances than those of a new person. This final alternative seems somewhat unlikely (especially given that a few of their items overlapped with ours), but could be ruled out by our replicating the current results using the Parise & Csibra items, or vice versa. A better understanding of this set of results awaits further research to test some of these proposed alternatives.

Another question raised by these data is what the mechanism underlying infants’ ability to recognize word-forms from a new speaker might be. In the adult literature, there is debate between traditional ‘abstractionist’ models on the one hand, which do not posit a role for idiosyncratic speech properties in word-form representation, and episodic models of speech on the other (e.g., MINERVA, Hintzman, 1986), which do include such information. Intermediary models allow for idiosyncratic speaker information to be represented in memory, but do not grant it a lexical role; this type of model seems to account well for existing adult data (see Cutler, Eisner, McQueen, & Norris, 2010 for a review).
The current results suggest the same type of intermediary model may account for infants’ performance. That is, the results of Experiment 1 suggest that that some form of abstraction has certainly taken place in infants’ lexical representations for food and body-part words, while other studies suggest that infants at this same age are able to detect idiosyncratic speaker information from previous ‘episodes’ (Houston & Jusczyk, 2000; Singh, et al., 2008, inter alia). Taken together, these results suggest that infants’ word-form representations are subject to the same mechanisms and representations as adults’: while an abstract form is stored in the lexicon, speaker-specific information may be retained to some degree as well.

Thus, further research is needed to investigate the mechanisms underlying our findings, and more generally, the processes underlying speech-invariance in infancy. It will be especially informative to determine balance of abstract and indexical information in infants’ word-representations. In this first step towards understanding infants’ speech invariance, we have shown that infants’ are able to generalize across idiosyncratic speaker differences during early word comprehension. We next ask whether infants generalize appropriately across phonemic differences during word comprehension as well.

**Experiment 2**

In Experiment 2, we investigated the precision of infants’ word-form representations. More specifically, here we queried whether infants treat mispronounced words as a poor match for portrayed referents. Previous research in this area has fallen into three broad categories: discrimination of recently heard speech stimuli, new word-new referent learning, and known word comprehension. Work with younger infants falls into the first category, typically examining infants’ sensitivity to recently trained lexical items, in the absence of a referential world, under the assumption that young infants understand very few words. For example, Jusczyk & Aslin (1995) find that by 7.5 months, infants recognize a word they were familiarized to, but not if the word is mispronounced, e.g. infants do not mistake ‘gike’ as an instance of ‘bike’.

The latter two categories typically consist of research with older infants. In the new word-new referent learning studies, infants’ ability to learn new words that sound similar is queried (e.g. Nazzi, 2005; Stager & Werker, 1997). In the domain of consonants, findings suggest that around 14 months infants may have trouble learning similar sounding labels (e.g. ‘bih’ vs. ‘dih’), depending on the task conditions, while younger and older infants succeed (Fennell & Werker, 2003, 2004; Rost & McMurray, 2009, 2010; Stager & Werker, 1997; Yoshida, et al., 2009). In the domain of vowels, Nazzi et al (2005) find that 20 month olds find it difficult to learn novel names for objects that differ only in vowel quality (e.g. pizeh vs. pizoo) in a fairly challenging referent-selection task.

In known word comprehension, the results have differed from the novel word learning tasks, most likely due to the longer training period, namely, infants’ entire life experience, and the nature of the task, i.e. recognizing referents for known words, rather than forging new label-referent mappings. Here, research has shown that from 12 months onward, infants are sensitive to vowel and consonant mispronunciations to very similar
While there has been debate over whether one might expect or predict differences in infants’ sensitivity to consonants and vowels during word-learning or comprehension (Havy & Nazzi, 2009; Hochmann, Benavides-Varela, Nespor, & Mehler, 2011; Nespor, Pena, & Mehler, 2003; Pons & Toro, 2010), here we choose to manipulate vowels because they are suggested to narrow in infants’ native language phoneme development earlier than consonants, by 6 months (Kuhl, et al., 1992b; Polka & Werker, 1994) rather than by about 12 months of age (Werker & Tees, 1984).

Another debate in this literature has concerned the nature of the mechanism underlying word-form specificity. One set of views, dubbed the developmental account, suggests that infants’ early representations are vague, and that as their vocabulary grows in size, infants are forced to add further detail to their representations to tell words apart. That is, learning more words leads to learning words that sound more similar (i.e. the minimal pair ball-doll), which in turn forces infants to learn the phonetic details of these words (Charles-Luce & Luce, 1990; Macnamara, 1982). An alternative view suggests that infants’ familiarity with a word dictates its degree of phonetic specificity, independently of overall vocabulary size (familiarity account).

When these views are tested using a mispronunciation paradigm (e.g. when infants hear ‘tog’ when seeing, e.g., a dog and a ball), the predominant result is that there is no relation between infants’ performance in mispronunciation tasks and their receptive or productive vocabularies (Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Mani & Plunkett, 2007; Swingley & Aslin, 2000, 2002). While a few studies have found a vocabulary-mispronunciation link, the data supporting this link are somewhat weaker. Namely, rather than a straightforward correlation between comprehension vocabulary and size of mispronunciation effect, these studies find correlations for only a subset of vocabulary measures, a subset of vocabulary sizes, and/or a subset of types of phonetic changes (Mani & Plunkett, 2010; Werker, et al., 2002). This in turn suggests that the more parsimonious of the two accounts is the familiarity account, though some caveats to its global applicability across paradigms, ages, and measures may apply. In the present study we collected vocabulary checklists from parents using MacArthur Communication Development Inventory (MCDI, Dale & Fenson, 1996), in order to offer evidence from younger infants concerning this ongoing debate.

Thus, building on the previous literature, Experiment 2 queries whether much younger infants just at the cusp of understanding words, i.e. 6-8 month olds, are able to detect a mispronunciation in a word that, when correctly pronounced, they comprehend. Given that we wanted to create the most direct comparison to Experiment 1 and B&S12, we chose to alter the main stressed vowel in each word used in that study, and also include an older group, 12-14 month olds, for comparison with B&S12 and other studies in the literature. We predict the older group will show decreased comprehension of the mispronounced words, in keeping with previous research (Mani & Plunkett, 2010; Swingley & Aslin, 2000, 2002). It is harder to predict what will happen in the case of the younger infants. If infants’ discrimination skills at this age (as demonstrated in Jusczyk & Aslin, 1995) are easily carried over and applied during word comprehension, we predict
infants will show decreased performance on the mispronounced words as compared to B&S12. On the other hand, if early word comprehension is dependent on a less-well specified lexical representation, or one that can be triggered by imperfect input than the representations drawn upon for speech discrimination, we would predict that infants will show above-chance comprehension, as in B&S12.

Methods

Participants

Subjects were two groups of infants, aged 6-8 months (N=17, M=7.22, R=6.04-8.99), and 12-14 months (N=11, M=13.58, R=12.16-14.91). Infants were recruited according to the same criteria and from the Philadelphia area as in Experiment 1 above.

<table>
<thead>
<tr>
<th>Original Word</th>
<th>MP</th>
</tr>
</thead>
<tbody>
<tr>
<td>apple</td>
<td>opal</td>
</tr>
<tr>
<td>banana</td>
<td>banoona</td>
</tr>
<tr>
<td>bottle</td>
<td>biddle</td>
</tr>
<tr>
<td>cookie</td>
<td>khaki</td>
</tr>
<tr>
<td>juice</td>
<td>jouse</td>
</tr>
<tr>
<td>milk</td>
<td>mulk</td>
</tr>
<tr>
<td>spoon</td>
<td>spoan</td>
</tr>
<tr>
<td>yogurt</td>
<td>yaygurt</td>
</tr>
<tr>
<td>ear</td>
<td>or</td>
</tr>
<tr>
<td>eyes</td>
<td>ayes</td>
</tr>
<tr>
<td>face</td>
<td>fouse</td>
</tr>
<tr>
<td>foot/feet</td>
<td>float</td>
</tr>
<tr>
<td>hair</td>
<td>har</td>
</tr>
<tr>
<td>hand/hands</td>
<td>hund/s</td>
</tr>
<tr>
<td>leg/legs</td>
<td>loog/s</td>
</tr>
<tr>
<td>mouth</td>
<td>mith</td>
</tr>
<tr>
<td>nose</td>
<td>nazz</td>
</tr>
</tbody>
</table>

Table 1: Stimuli used in Experiment 2. On the left, ‘Original Word’ indicates the word as pronounced in B&S12, and Experiment 1. On the right, ‘MP’ (Mispronunciation) reflects the pronunciation of that same word in Experiment 2.

Materials

The visual materials were identical to those presented in B&S12. The auditory materials varied from those in the original study in that each noun was deliberately mispronounced, with a change in the stressed vowel. The vowel shift was selected with several criteria in mind: it should be maximally different from the original vowel, not
create a real word that would be known to infants, and not be too similar to the vowel in any of the competitor words displayed simultaneously with that target. See table 1.

**Apparatus and procedure.**

The apparatus and procedure was identical to that in B&S12, except in the change of materials, described above, to which parents were familiarized. In order to familiarize parents to the mispronounced (MP) words they would be uttering, parents heard the words they would be repeating over headphones before the experiment began. While hearing them over headphones, they also saw them in a list written out phonetically, along with a ‘sounds like’ or ‘rhymes with’ aid, e.g. for the MP of ‘nose’ they saw ‘nazz, rhymes with jazz’. Parents did not see the MP next to the word it was an MP of, in order to help prevent parents from accidentally saying the correctly pronounced form during the experiment.

**Results**

**Experiment 2, paired-picture trials.**

Performance on the paired-picture trials differed by age-group (see Figure 2, right panel). In the younger group, collapsing over subjects, 12/17 6-8 month olds looked longer at the correct image, resulting group performance that did not differ from chance (M=.051, p=.098). Over item-pairs, 6-8 month olds showed positive difference scores on 6/8 item-pairs; this too did not differ significantly from chance performance (m=.043, p=.10).

In the older group, over subjects, and over item-pairs, the 12-14 month old sample did not perform above chance. 5/11 12-14 month olds showed positive difference scores (m=.0054, p=.58). Over item-pairs, 12-14 month olds showed positive difference scores on 6/8 item-pairs (m=.018, p=.37).

**Experiment 2, scene trials.**

Performance over the younger group of subjects in the scene trials in this study was above chance; this is somewhat anomalous, and perhaps explained by the reduced sample size in this trial type, for which fewer infants met the inclusion criteria. 6/8 6-8 month olds showed positive corrected proportions of target looking (M=.053, p=.020). Over items, 6-8 month olds showed positive performance on 9/16 items, which did not reach significance (M=.042, p=.091).

For the scene trial performance in the older group, 3/11 12-14 month olds showed positive corrected proportions of target looking (M=-.027, p=.84). Over items, 12-14 month olds showed positive performance on 5/16 items (M=-.023, p=.88).

**Age group comparison, experiment 2.**

We then asked whether the two age groups in this study differed from each other. For paired-picture trials, 6-8 and 12-14 month olds performance did not differ by two-tailed Wilcoxon test (estimated difference=.052, p=.38). For scene trials, 6-8 and 12-14 month olds performance differed marginally (estimated difference=.098, p=.051).
Pair-wise comparison with B&S12, and with experiment 1.

Showing that infants perform at chance when hearing their parent mispronounce the target image’s label still leaves open the possibility that they nevertheless showed performance that did not vary from the correct pronunciation. We thus compared performance in the original B&S12 data set with the present study’s data (see Figure 2). In a two-sample comparison by two-tailed Wilcoxon test, 6-8 month olds’ data from the original study and Experiment 2 did not differ for either trial type (paired-picture trials: estimated difference=.038, p=.38; scene trials: estimated difference=.064, p=.15), while 12-14 month olds’ data across the two studies did, with performance by 12-14 month olds significantly worse in Experiment 2 for paired-picture trials, and marginally worse for scene trials (paired-picture trials: estimated difference=.13, p=.039; scene trials: estimated difference=.074, p=.095).

We next compared performance in the younger group in Experiment 1 and 2 (but did not compare older group performance since the older group across experiments was of a different age-range). For paired-picture trials, performance by the 6-8 month olds in Experiment 1 and 2 did not differ (estimated difference=-.004, p=.97); For scene trials, performance was better in Experiment 2 (estimated difference=-.074, p=.039).

Modeling results for experiment 1, 2, and B&S12.

We next sought to compare 6-8 month olds’ performance across Experiment 1, Experiment 2, and B&S12. We modeled the trial-types separately since different outcome measures were used across trial-types. We entered infants’ difference scores for each item-pair from the paired-picture trial-type into a multi-level mixed effects model using Subject and Item-Pair as random effects, and Study as a fixed effect, with B&S12 as the reference level of this predictor. This model finds that the intercept is significantly different than chance (estimate=.067, T=2.574) but that Study is not a reliable predictor of infants’ performance (F=.46, on 2 degrees of freedom); that is, neither Experiment 1 nor Experiment 2 differed significantly from B&S12 in this model. Indeed, a model omitting study is not significantly worse than one that includes it (2*log-likelihood chi square=.92, df=2, p=.63). Models including age, gender, and mother’s socioeconomic status were no better than models without these variables. These modeling results confirm the pair-wise comparisons reported above: the six-eight month data did not vary across the three studies. Analogous models for the scene trial-type showed the same pattern: neither experiment differed significantly from B&S12 (F=1.9 on 2 degrees of freedom), and the model including study was no better than one omitting it (2*log-likelihood chi square=.38, df=2, p=.15)

Next, we modeled performance across these three studies including subjects of all ages. Again, for paired-picture trials we used infants’ item-pair level data in a multi-level mixed effects model using Subject and Item-Pair as random effects, and Study as a fixed effect, with B&S12 as the reference level for this predictor. Here we find that a model including only Study as a predictor finds main effect of study (F=2.51 on 2 degrees of

15 Similar models of the scene-trial data also found that Study was not a significant predictor.
freedom) and that performance in Experiment 2 is significantly worse than in B&S12 (T=-2.092; estimate=-.061), but Experiment 1 performance does not differ from B&S12 (T=-1.53, estimate=-.037). A model including the interaction of age and study finds a significant effect of the interaction between Experiment 2 and Age (estimate=-.021, T=-2.191), but no other significant effects. However, models including age do not provide a better fit to the data than those that omit it, whether it is included as an interaction with study or as a simple effect. Due to the large number of missing data-points, which are not randomly distributed across ages, and the smaller number of subjects retained in the scene trial analyses, the analogous scene-trial models were not run (see Appendix A).

Discussion

In Experiment 2, we found that our two age-groups were differentially affected by a vowel mispronunciation in the target word. Performance in the 6-8 month old group was mixed. For both trial-types performance did not differ from the older infants’ performance, but it also did not differ from performance in B&S12. For paired-picture trials, infants failed to look at the target image at above-chance rates. For scene trials, performance was above-chance (though this is likely a statistical anomaly, due to the small number of subjects in this subset (n=8); see Appendix A).

In contrast, the pattern of results with older infants was much more clear: by 12-14 months of age, mispronounced target words lead infants to fail to look at the correct image at above-chance rates. Indeed, their performance, for both trial-types, was significantly worse in Experiment 2 than in B&S12, showing that a single changed vowel negatively impacts infants’ word comprehension.

These results suggest that the first step for following up on Experiment 2 is to increase the sample size in the younger age group, to determine whether the mixed pattern of results is perhaps due to small effects that are not found without greater statistical power, or whether indeed performance at this age is mixed. The next step is to extend the study to an intermediary age-group, so that the full 6-14 month range can be compared to B&S12, and to Experiment 1.

Kind and size of change.

Experiment 2 made use of a salient phonemic change: the stressed vowel of each target word was changed to another vowel that differed maximally within the constraints of the experiment (see Table 1 and Materials, above). This was in part to establish a baseline condition in which infants were most likely to notice the phonemic change, given previous literature showing early native-language vowel sensitivity (Kuhl, et al., 1992b; Polka & Werker, 1994).

With this vowel change, 12-14 month olds showed clearly degraded performance compared to B&S12, replicating the findings of others (e.g. Mani & Plunkett, 2010; Swingley & Aslin, 2000). In contrast, 6-8 month olds, who have not been previously tested using a word comprehension paradigm to query word-form representations in this way, showed a mixed pattern of performance. Thus, an open question for future work is to better understand the development of word-form specificity between 6 and 12 months, and to begin to test other kinds of phonemic contrasts, e.g. by varying a consonant instead
of a vowel, by systematically varying a given number of features, in a given position, in words of a controlled length, etc. (e.g. Mani & Plunkett, 2010; Nazzi, 2005). Since previous research has found that around their first birthday, infants are equally sensitive to mispronunciations of varying sizes (e.g. ‘tog’ and ‘bog’ for ‘dog’, Swingley & Aslin, 2002), and equally sensitive to certain consonant and vowel mispronunciations (e.g. ‘kep’ and ‘tup’ for ‘cup’, Mani & Plunkett, 2010), it may be especially interesting to look at these types of changes with younger infants to see whether infants’ earliest word comprehension is equally affected by these types of phonetic changes as well.

**Vocabulary and word-form precision.**

As discussed above, there is debate over the nature of the relationship between vocabulary size and word-form precision. The developmental view holds that as the size of the vocabulary increases, the precision in stored lexical forms increases so that similar sounding words can be kept distinct, while the familiarity view holds that the amount of experience with a given word, regardless of vocabulary size, is what determines the degree of stored precision. In previous research, the familiarity view has received more empirical support, though this has only been queried using mispronunciation tasks such as the one used here in infants 12 months or older (Bailey & Plunkett, 2002; Swingley & Aslin, 2000, 2002).

Using the MCDI vocabulary data provided by parents, we can consider our data in light of this debate. Given that our two groups of subjects varied in age, and in their pattern of performance in relation to B&S12, we looked for correlations between vocabulary size and task performance separately for each age group. Indeed, unsurprisingly, and in keeping with our findings in B&S12, the younger infants had much smaller reported vocabularies than older infants (6-8s month olds, median comprehension vocabulary=3 (R=0-49); median productive vocabulary=0 (R=0-2); 12-14 month olds, median comprehension vocabulary=77 (R=23-217); median productive vocabulary=11 (R=2-29)). In principal, one might argue that younger infants having a smaller vocabulary and a mixed pattern of performance with the mispronunciation task provides support for the developmental account, but to really evaluate this, it is necessary to look at the correlations between infants’ performance and their vocabulary size within each age group.

We computed nonparametric correlations between vocabulary size and performance in the eyetracking task. Using four different (but related) measures of vocabulary size (overall productive vocabulary on the CDI, overall comprehension vocabulary on the CDI, number of target words reportedly said on the CDI, and number of target words reportedly understood on the CDI), we found no evidence for a relationship between these vocabulary measures and task performance for either age group. All correlations were weak and none reached statistical significance (Kendall’s tau between -.3 and .1; all p>.2).

---

16 We use the paired-picture data here, given that this condition is more analogous to the previous literature
Thus, our data are more compatible with the familiarity account than the developmental account, and suggest that infants’ word-form specificity is unrelated to their vocabulary size. Given younger infants smaller vocabularies, one could imagine a better test of this question using a longitudinal design; this is open for future research.

**General Discussion**

Taken together, Experiment 1 and 2 increase our understanding of infants’ early word-form representations. Experiment 1 shows that infants have one type of generalization available to them very early on: the ability to recognize words across speakers. This study leaves ambiguous when this ability develops. It is possible that infants are able to map incoming speech into a normalized acoustic space innately, at least for certain indexical properties. On the other hand, it is possible that this ability must be learned, in which case our data suggest that it is learned to some approximation before infants reach six months of age. Experiment 2 shows that another kind of generalization across word-forms, namely, recognition of words only when they are pronounced with the proper phonemes, may develop between 6 and 12 months of age. That is, while the data from Experiment 2 indicate that by 12 months infants appropriately make this generalization, replicating previous work (Mani & Plunkett, 2010; Swingley & Aslin, 2000, 2002), the data at 6-8 months are less clear, suggesting the possibility of a change in word-form specificity between these age-groups. Thus, future research is needed to clarify and extend our understanding of the developmental timecourse of each of these kinds of generalizations.

The paradigm employed here, namely, examining word-form representations through a word-comprehension task, is methodologically very different from the more standard discrimination-based paradigms typically used to look at this issue (e.g. familiarization followed by the head-turn preference or conditioned-head-turn paradigms). The latter generally expose infants to a set of speech stimuli, and then after this familiarization, play other sets of stimuli that are related in certain ways to the familiarization, to see which set of new stimuli infants prefer. This is generally done in the absence of referential displays, and often with words infants do not know. These paradigms are very powerful in that they allow researchers to test a wide range of phonological questions with young infants, but are limited in the specific conclusions that can be drawn from them.

Research within these frameworks has shown that in the first year, infants are sensitive to various indexical properties, such as changes in gender or affect (Houston & Jusczyk, 2000; Singh, et al., 2004; van Heugten & Johnson, 2012), as well as phoneme changes (Jusczyk and Aslin, 1995). This work is notable for showing that infants do indeed store abstract information about both kinds of properties examined in the present research, that is, indexical properties and phonetic structure of word-forms. However, as the studies above demonstrate, discrimination of, e.g. indexical properties does not imply differentiated word comprehension in light of indexical differences. That is, discrimination studies leave unclear how infants *use* their discrimination abilities during linguistic interactions, e.g. during word comprehension, or more broadly, in the language acquisition process. While it is difficult to compare these discrimination-based findings
with our own directly, it will be important for future work to determine what the timelines determined from both of these types of experiments tell us about how linguistic development proceeds.

In conclusion, in the studies presented above, we find that indexical differences, e.g. those between voices, do not prevent infants from understanding known words, at the earliest age those words have been shown to be understood, i.e., 6 months. In contrast, by 12 months and possible earlier, infants comprehension is impaired if a phoneme in these same words is changed. These results suggest that infants’ early word-form representations follow the same kinds of generalizations as those of adults: non-phonemic changes do not impair comprehension, while phonemic changes do. Thus, young infants not only appropriately recognize, but also understand words they know in the right set of acoustic contexts. This is made all the more impressive given that these infants by and large do not yet say the word-forms that they’re hearing, storing, accessing, understanding, and generalizing over in adult-like ways.
Appendix A: Scene Trial Discussion

As mentioned above, the data from the scene trials for both 6-8 and 9-12 month olds were not as strong as the data from the paired-picture trials. In B&S12, these trials also showed fragile performance: infants succeeded at above-chance rates as a group of 6-9 month olds, and in the 6-7 month old subset, over subjects, but not in the group of 8-9 month olds nor in any age subset in the items-analysis. There are many possible reasons for this, discussed below. While the failure of infants in these trials in Experiment 1 is not problematic for our comparison with B&S12 (see main text), a more detailed discussion of these trials is still useful for a better understanding of our data, and what it implies both theoretically and methodologically for future studies.

First, on theoretical grounds, it may be the case that picking out a referent among several competitors in the same semantic category is simply harder to do than picking out a referent compared to a single competitor from another semantic domain. That is, infants may know that a ‘mouth’ is not an ‘apple’, but not that it’s not a ‘nose’, or even if they do know what the referent looks like with some specificity, seeing more complicated images may make it harder for them to find the part of the picture they are looking for very easily. This issue can be addressed by future experiments, which could vary number of competitors, and their degree of semantic relatedness parametrically.

Several other reasons for infants’ poor performance on these trials concern problems that are not relevant to their linguistic abilities, but rather to the nature of the stimuli and analysis problems for such stimuli. For instance, as discussed in B&S12, in this age range, infants are showing increased attention to faces and eyes in particular, and thus those images garner more attention regardless of the linguistic stimulus. Moreover, even among the food images, some parts of the image draw attention more robustly than others, depending on how the items in the image are arranged, and what they are (e.g. bottles may be more interesting than spoons to infants, all else equal). Young infants in particular have difficulty disengaging from highly salient stimuli (Frank, Vul, & Johnson, 2009).

Another problem with the scene trials lies in how to analyze them. The standard analysis involves drawing areas of interest around parts of the images and asking what proportion of the time infants fixate those areas over a given time-window. This is potentially problematic for several reasons. First, proportion-of-target-looking analysis collapses a rich data set into a single proportion for each trial or item, though this is somewhat unavoidable given that the time-course analyses used with adults are less viable with infants’ much noisier data; this point is general to all infant eye-tracking research that uses such analyses. Second, in the scene trials, unlike the paired-picture trials, the areas of interest were of varying sizes given the real-life differences in size in, e.g. a spoon and a banana, or hands and legs. Third, in determining whether the proportion of looking to a target region is significant requires a comparison to ‘chance’, which is poorly defined here. With two equally sized images, as in the paired-picture trials, chance is 50%; it is much less easily quantified with interest areas of varying sizes, and all the more so for body/face images where the interest areas do not have clear boundaries.
One way to try to solve both the saliency and analysis issues mentioned above that is common in the field is to use a ‘baseline’ proportion from the period of time before the target is labeled, that is then subtracted away from the proportion of target looking in the window of interest; this is the approach we take here. While some sort of correction along these lines is likely appropriate and even necessary with young infants’ data, it is not clear what the nature of the relationship between the baseline and post-target period should be. The type of correction, implemented by us here as by others, may over-correct for high-saliency parts of the image. That is, if infants are already looking at the eyes 95% of the time, there is not much room for them to show increased looking after the target is named. Moreover, performing this correction requires infants to have been looking for a sufficient amount of time before the image was named, which, with fidgety infants, results in fewer trials entering the analysis. Along similar lines, the ‘difference-score’ analysis used for the paired-picture trials, in which we calculate how much more infants look at an image when it’s the target than when it’s the distracter, resulting in pair-level data (see B&S12 for more details) is much harder to implement for the scene trials: it requires that infants contribute sufficient data (i.e. look for more than 1/3 of the window of interest, and look at more than one part of the image) on all four trials for a given scene image. This was often not the case, resulting in a very large number of missing cells.

Finally, while this seems to paint a dire picture of these trials, we should point out that these problems disproportionately affect the younger infants: infants over 13 months showed highly robust performance on this trial-type in B&S12 (as they did on the paired-picture trials). This too could be for many possible reasons, but certainly is at least partly explained by older infants’ increased ability to disengage from salient stimuli, and to attend to the experiment to a greater degree, as well as by their better control over various aspects of language.

In conclusion, given that this is one of the first eye-tracking studies with young infants examining word comprehension using complex images (though cf. Aslin 2009 for a non-linguistic task), it is perhaps not altogether unexpected that many methodological issues have arisen. Given that infants day-to-day life is more similar to the complex scene images than the sparse paired-trials, it will be important for future work to further investigate infants’ word comprehension during complex image viewing, with the issues raised here in mind.
Chapter 5: Conclusions, Further Directions, and Discussion of
Word-Learning Mechanisms

Summary of Dissertation

In this dissertation, I have demonstrated several new findings about young infants’ word learning, taking the first steps towards answering the questions of when infants first comprehend words in their native language, and what kinds of words they learn initially. First, by around 6 months of age, infants are able to understand around a dozen words for foods and body-parts, whether these words are labeled by their mother or by an unfamiliar experimenter (Chapter 2 & 4). Second, by around 10 months of age (but no sooner), infants’ lexicon has expanded to include more abstract non-nouns such as ‘hi’ and ‘all gone’ (Chapter 3). Both of these findings place comprehension of these groups of words approximately a full half-year earlier than was assumed or had been demonstrated in the literature (e.g., Bloom, 2002; Kuhl, 2011), with the notable exception of Tincoff and Jusczyk (1999, 2012) for ‘mommy,’ ‘daddy,’ ‘hand,’ and ‘foot’. Third, spanning these time periods, we find that infants’ phonetic precision undergoes a shift between 6 and 14 months: at 6 months they show some signs of sensitivity to the precise phonemes in words they understand, a sensitivity which increases by their first birthday; by 12-14 months there is a significant difference between their comprehension of correctly and incorrectly pronounced words. Fourth, for both nouns and non-nouns, infants’ comprehension abilities are at first numerically modest, though robust across infants and item-pairs; around 13-14 months, infants’ performance in our studies increases greatly to reflect consistently high levels of comprehension.

These findings raise several issues which merit further discussion, many of which can only be resolved through future research. While this dissertation has dealt mainly in establishing the ‘what’ and ‘when’ of early word-learning, the results have opened up many questions about the ‘why’ and ‘how’ of word-learning. Although we proposed some ideas concerning these more challenging questions in the preceding chapters, both theoretical and experimental advances are necessary to create a more unified account of word learning between 6 and 16 months. As explained in the introduction, before this age, word comprehension is yet to be demonstrated, and after this age, infants have other linguistic abilities at their disposal, so this is the particular age range of greatest interest for early word-learning. Below, I summarize and build on ideas concerning the relevant developments we have uncovered in the preceding chapters, and describe ongoing work or propose future work that can help adjudicate between various hypotheses on the table.

Given the timelines this work has charted, it is important to understand linguistic, and more broadly, cognitive and social developments within these timeframes, and to examine how these developments might link to word comprehension. First, what happens before 6 months to allow infants to learn food and body part words, regardless of speaker, and with still developing phoneme inventories? Second, what changes between 6-10 months that allows infants to understand their first non-nouns, which do not, for instance, have a core set of visual features in the same sense the nouns do? Third, what underlies the shift in infants’ word-form precision between 6 and 12 months? Finally, what
happens between 10 and 13 months that leads to the dramatic improvement in word comprehension that we find at that time? I address these issues below, and provide discussion of current and future work that could help us better understand these transitions. Then I turn to some larger questions about word learning, and its mechanistic underpinnings.

**Birth to 6 months**

While a lot of space was devoted to describing infants’ early linguistic and conceptual knowledge in Chapter 1, it is useful to reevaluate this work in light of our findings that infants already know around a dozen words for foods and body-parts by 6 months, and for more abstract non-nouns about four months thereafter. Thus, even though it is clear that many aspects of language acquisition and conceptual development continue to improve over the first year and beyond, there must be some minimal scaffolding upon which infants build their word-referent links before infants reach half a year of age. In this light, findings of infants’ abilities before six months are especially potent prerequisites; skills attained around the same time that babies demonstrate word comprehension may or may not have been used in learning the queried words. However, given that word comprehension, or, indeed, most abilities found in 6 month olds likely do not become available to each infant exactly on her half birthday, we consider here evidence of other abilities at 6 months as potentially recruited in word comprehension as well.

**Linguistic skills before six months.**

By six months, infants’ acquisition of their native language phonemes is underway. While infants’ consonant inventory is still narrowing, their vowel space is already settling into its native-language organization (Kuhl, et al., 1992b; Polka & Werker, 1994). However, it is not entirely clear what the basis of infants’ early word-form representations are, i.e. whether their initial representations are adult-like, including both phonemic and indexical information to some degree, or whether initial representations are more closely tied to acoustically-based discrimination abilities that over time develop into phonemic and indexical knowledge. As the results of Chapter 4 demonstrated, infants’ initial lexical representations show generalization over certain indexical properties by 6 months. Moreover, by 6 months infants prefer speakers of their native language (Kinzler, Emmanuel Dupoux, & Spelke, 2007), suggesting not only a discrimination of their native language’s sounds, which is found in newborns (Mehler, et al., 1988), but a related social preference predicated upon this discrimination by 6 months.

In terms of word-form recognition, our results suggest that infants are able to recognize words and link them to their referents by around 6 months (Chapter 1). Given the assumption that word segmentation is necessary for word recognition, this may seem to present a curious contrast to the results of Jusczyk & Aslin (1995), in which 7.5 but not 6 month olds listened longer to passages containing words they had heard in isolation. This in turn suggests that word segmentation is not available before 7.5 months, raising
several possibilities about how our infants may have succeeded in segmenting out our target words from their daily input.

First, it could be the case that word segmentation as operationalized by an exposure to somewhat low-frequency words followed by tests of the words in passages may not reflect word segmentation with more frequently heard words. That is, the failure of 6 month olds could reflect the methodology or item-choice of Jusczyk and Aslin (1995), rather than 6 month olds’ abilities more globally.

Another possibility is that the methodology is not at fault, but that 6 month olds can segment words only in certain situations. This option is supported by the work of Boruff et al. (2005), who find that 6 month olds can indeed recognize a word they were familiarized to in a passage, if it follows their name or their appellation for their mother. Another potential way 6 month olds could be segmenting words is by beginning with words that occur commonly sentence-initially or in isolation. That is, perhaps one reason for 6 month olds’ failure in Jusczyk and Aslin (1995) is that words were only at an edge (i.e. beginning or end of an utterance) in a third of the sentences, and then only sentence-finally; as Seidl and Johnson have demonstrated (Seidl & Johnson, 2006), even 7.5 month olds have an easier time extracting words from edges.

Another possibility is that infants could succeed in a word segmentation task at 6 months if it were paired with visual input: perhaps having potential referents to tie repeating word-forms to gives infants traction in forming a representation of new words. This option finds some support from Shukla et al.’s demonstration of word learning by 6 month old infants shown a looming referent paired with artificial language exposure (2011, see Chapter 1 for further detail).

Even more compelling, recent research (Seidl, Tincoff, Baker, & Cristia, submitted) highlights one mechanism infants may recruit for segmenting body-part words in particular. These authors find that 4 month olds use tactile cues presented in synchrony with a stream of artificial language (which was stripped of all other cues to segmentation) to help them pull words out of the speech stream. More concretely, infants who received a consistent touch to their elbow or knee during a given trisyllabic sequence of the artificial speech stream subsequently discriminated between this ‘always touch’ word and two other words (a ‘one touch’ word and a ‘nonword’.) Moreover, ruling out a more social-engagement based interpretation in favor of one specific to infants’ tactile experience, infants only showed this learning when their own body-part was touched, not when they saw the experimenter touch her body part while the trisyllabic word occurred. This finding provides an intriguing possibility that one way that infants learn body parts so precociously is through their incorporation of tactile and auditory information.

While studies in the literature suggest that infants are able to segment the speech stream and recognize word-forms in it by 6 months, and even earlier with tactile cues, further research is needed to establish the details of how this link occurs in real-life word learning. It could turn out that the segmentation problem for first-learned words is not particularly hard: perhaps first-learned words occur in isolation or at edges, with highlighted prosodic marking, or coupled with clear multimodal feedback often enough they can be detected by infants more easily, obviating the need for very robust word
segmentation to be in place before word learning begins. Analyses of video corpora from young infants could help determine the viability of this account.

Our findings gel well with those showing that at six months, infants can not only segment the speech stream, but can learn words as well. From brief, in-lab exposure they are able to learn novel object labels (e.g., Gogate, et al., 2006; Shukla, et al., 2011), and from real-life exposure they are able to learn proper names (Mandel, et al., 1995; Tincoff & Jusczyk, 1999), and the words ‘feet’ and ‘hands’ (Tincoff & Jusczyk, 2012). However, it is possible that these first two sets of studies recruit only partially overlapping mechanisms with Tincoff and Jusczyk (2012) and our own work. That is, word learning in each of these cases requires picking out words and potential referents to link together, perhaps through multi-modal cues, but learning common nouns may require more.

Proper names are in a unique position of picking out individuals, as opposed to common nouns, which pick out categories (Macnamara, 1982; Waxman & Gelman, 2009). While this is clear in the case of infants’ own name and appellations for her parents, it is perhaps more opaque in the case of novel object labels. However, while adults may assume that novel object labels refer to entire categories, the exposure given to infants in these studies provides no grounds for this assumption, as only one instance of the novel object is used. It may turn out to be trivial for infants to generalize known labels to other category members, but this is yet to be demonstrated. Relatedly, it is also unclear how infants might learn when to generalize labels to novel instances, and when not to; undoubtedly learning that other women are called ‘mommy’ is a shock for young children.

**Conceptual skills before six months.**

As we saw in Chapter 1, infants have many skills involving physical and social reasoning in place before 6 months. More mature skills, especially those concerning joint attention, seem to appear later, which suggests that these skills are not necessary to learn words for foods and body parts, but may be necessary to learn words for non-nouns.

Even though foods and body parts are not objects per se (see further discussion below) many of the same physical properties pertain. Thus, infants’ demonstrations of object permanence by 3 months (Aguiar & Baillargeon, 1999), and knowledge of other properties of object-hood, e.g. continuity of motion, by 4 months (Spelke, et al., 1995) could be recruited in their learning about early words. For instance, understanding objects’ general physical properties, e.g. existing when out of view and moving together, may allow infants to posit that such entities have consistent labels.

By 6 months, infants also have the rudiments of probabilistic reasoning (Denison, Reed, & Xu, 2013). One could imagine that the same reasoning underlying infants’ surprise when blindly selected items do not match those items’ distribution in a seen population could be harnessed to match words with likely referents. That is, infants may weigh the probability of a word-referent link, given the population of evidence, i.e. infants’ experiences in the world.

In the social domain, infants attribute preferences to agents as early as 3-5 months (Luo, 2011; Woodward, 1998, 1999). By 6 months they have refined this knowledge a bit, showing an understanding of how another’s perception constrains her preferences
(e.g. whether an experimenter can see both objects before expressing preference for one of them) (Luo & Johnson, 2009). At this age too, perhaps relatedly, infants begin to show a rudimentary ability to follow another’s gaze (Morales, et al., 1998). Of course these skills continue to become adult-like over the coming years, but are in place in a measurable way before 6 months. This development suggests that an understanding of preferences, and even a rudimentary ability to follow gaze may be useful (or even a prerequisite) for understanding words.

**Foods and body parts, not Spelke-objects.**

One finding in this dissertation concerns the types of words infants are showing early knowledge of, namely foods and body parts. This is perhaps unexpected given that infants’ knowledge of and physical reasoning about bounded, standalone objects precedes their understanding of liquids, shape-shifting substances, and foods, which are sometimes tied to language-specific classifiers, and in any event are not demonstrated to be well understood until later in infancy (Imai & Gentner, 1997; Shutts, Condry, et al., 2009; Soja, Carey, & Spelke, 1991).

At the same time, perhaps it not entirely surprising that these are the words that infants begin with: indeed they are very salient aspects of infants’ everyday lives. Infants are bathed, changed, dressed, and fed over the course of the day, leading parents to very frequently use food and body part words, as determined in a corpus of mothers interacting with their infants in the home (Brent & Siskind, 2001). Moreover, these items, while not bona fide objects, do tend to have the same shape across instances and across time, and have object-like attributes. That is, even though a hand can be balled up, ranges across shades of brown, and does not stand alone (except in The Addams Family), it still has a core set of physical properties. So too in the case of juice, which takes on the shape of its container. In infants’ experience, however, these containers are surely limited in shape and size, mostly to bottles and sippie cups. Thus, while recognizing the images in our studies did require the infants to generalize from their own experiences of their foods, body parts, etc. to our 2-dimensional images, we did not seek to push the limits of this generalization by giving atypical instances of our items for the infants to discriminate, and as such, these items were perhaps not so very different from concrete objects.

Along these lines, Bloom has discussed the acquisition of words for parts and actions, like ‘eye’ and ‘kiss’: “… while being a Spelke-object may be a sufficient condition for being a nameable individual, it is plainly not a necessary one” (Bloom, 2002, p. 108). Though he was writing in reference to 2 year olds learning these words, the same points apply here to infants’ knowledge. Bloom goes on to describe that words like body parts are psychologically natural and fairly close to objects, conceptually: while they do not show cohesion or independent movement to the degree that stand-alone objects do, body-parts generally move together, connectedly, could theoretically be severed, and can move independently to some degree (see Bloom, 2002, Chapter 4). Thus, this distinction of being a ‘psychologically natural part’ coupled with the frequent and salient experiences infants have with foods and body-parts, both visual, and tactile, is in line with our finding that these are learned early in development.
Summary of developments before six months.

In the months leading up to the time that we find knowledge of food and body-part words, infants have a fair number of abilities in the conceptual and linguistic domain. They know the vowels in their language, and use known words to help them find others (Bortfeld, et al., 2005; Kuhl, et al., 1992b; Polka & Werker, 1994). They understand some basic properties about objects and agents (Baillargeon, et al., in press; Morales, et al., 1998; Spelke, et al., 1995; Woodward, 1998, 1999). They are also able to form basic links between newly heard words and their referents; this last finding, perhaps aided by cross-modal redundancy (Gogate & Bahrick, 2001; Seidl, et al., submitted; Shukla, et al., 2011). While all of these abilities show growth over the following years, it seems that their early form is enough to allow infants to break into the word-comprehension system, and begin to learn the meanings of foods and body parts in their native language.

Between 6-9 and 10-13 months

Along with demonstrating that 10-13 month old infants understand non-nouns, like ‘all gone,’ Chapter 3 began to explore what may change between 6-7 months, when infants first understand food and body-part words, and 10 months, when they first understand early non-nouns. In Chapter 3, we proposed two sets of explanations for this timeline, summarized below, to which here I add a third type of potential explanation (drawn from Bergelson & Swingley, 2013a). These three accounts, the data-driven, social, and conceptual accounts, are not mutually exclusive, and indeed, current data suggest that elements of each of them likely underlie the delay in non-noun learning relative to noun-learning.

Data-driven account.

The data-driven account suggests that non-nouns are learned the same way as nouns, but that their learning is delayed because the statistics that underlie these words and meanings are messier. In our corpora analyses in Chapter 3 (Bergelson & Swingley, 2013b), we investigated two potential ways in which the data-driven account may manifest itself. Specifically, this chapter explored whether frequency and environment could explain the later onset of non-noun comprehension. The results of the two corpus studies in that chapter demonstrated that there is no difference in frequency between the nouns and non-nouns we tested: that is, infants are no more likely to hear the non-nouns than the nouns, either embedded in sentences, or in isolation, at least in the corpora we examined (Brent & Siskind, 2001; Demuth, et al., 2006).

In terms of how these words occur in infants’ environment, the analysis of the Providence video corpus found that on most variables, again, our group of nouns and non-nouns did not vary (e.g. in the settings they occur in, in whether the mother and infant are attending to the same thing, etc.). Where these words did differ was in how often parents said a word when its referent (used loosely as ‘instance of the word meaning’) was present. That is, when parents said the nouns, their referents were almost always visible in physical or picture form, or in the case of body parts, were especially focused on. In contrast, the non-nouns were much more likely to be said when the relevant referent or action was missing, i.e., fairly often, ‘bye’ was said when no-one was
leaving, ‘kiss’ was said when no kissing was occurring, etc. This difference, we suggest, may be part of the explanation for why these words are understood later by infants.

**Social skills account.**

The second account we propose is that learning non-nouns requires social skills that only emerge closer to 10 months, such as skills related to joint attention and gaze-following. Indeed, several studies have documented that better understanding of gaze and intention develops between 9 and 10 months (e.g. Beier & Spelke, 2012; Brooks & Meltzoff, 2005). These skills may be more necessary for words whose visual features are less consistent across instances, i.e. ‘uh-oh,’ as opposed to ‘banana.’

In an attempt to better understand the relation of social intention and word-comprehension, ongoing work in our lab examines infants’ understanding of gaze and pointing. We have recently found that indeed, around 10 months, infants’ ability to follow pointing increases in a way that seems qualitative (Bergelson & Swingley, 2013a); before this age infants are not very good at following an actress’s pointing towards a toy, whereas after this age they succeed in doing so. However, these social cognition findings are all merely suggestive of a link between social intent and non-noun learning, rather than indicative of a specific link. In a more direct test of this proposal, we are currently testing whether infants’ abilities to follow pointing correlate with their abilities to understand non-nouns in particular.

This account is in keeping with a proposal by Bloom (2002) that there are two distinct cognitive systems children use when learning words: a system for objects and a system for function and intent. The first system, through which we get objects and by extension parts, relies on physical reasoning, especially cohesion, laid out by Spelke, Baillargeon and colleagues; see chapter 1. The second system is Theory of Mind, which separates matter and motion through actors’ intentions and objects’ functions. Thus, to extend Bloom’s framework, the first system may come online earlier, and let infants learn nouns, while the second system’s coming online around 10-12 months allows for the learning of non-nouns.

**Conceptual account.**

One other account, discussed less explicitly in Chapter 3 (but see Bergelson & Swingley, 2013a), is the conceptual account. This account proposes that non-nouns are harder to learn because of the nature of the concepts and categories involved. Instances of a word like ‘all gone’ vary more and thus may be harder to recognize as having a common semantic core than instances of ‘hand.’ This hypothesis can be expressed as stemming from higher-level differences in the kinds of linguistic roles played by nouns in contrast with adjectives, exclamations, verbs, and social greetings. It can also be thought of as a low-level difference in what ‘features’ must be summed over: in the noun case, visual features such as shape, size, and color may be easily graspable from the environment, while non-nouns require more abstract (perhaps second-order) features, which may be harder to posit or grasp. A related hypothesis concerns biases in word-learning; it could be the case that in the absence of further evidence, infants choose to
posit that a new content word they have isolated from the speech stream refers to a noun before they consider that it may refer to another part of speech.

While one could imagine, in principle, several ways to test for this kind of noun bias, or noun conceptual advantage, in ongoing work in the lab we are examining this through a teaching study. The point of this study is to directly compare the learning of novel nouns and verbs, equalizing how often the referents appear when they are said, which, as we saw in Chapter 3, is one difference between nouns and non-nouns. In this study, infants take home four sets of toys, two that correspond to novel nouns, and two that correspond to novel verbs. Parents are asked to play with the toys with their infants, using the novel words, for ten minutes a day for two weeks. After this longitudinal at-home exposure, (which we monitor through audio recordings), infants come back to the lab for an eye-tracking experiment testing their learning of the new words in both simple and complex referential comparisons (i.e. we test whether they learn the word-object and word-action links using the specific materials they took home, and whether they extend the words to novel instances in an adult-like way). The results of this study may speak to the validity of the conceptual hypothesis in explaining the décalage between noun and non-noun word-comprehension of real words of English that we find in Chapters 2-4.

Summary of 6-10 month change accounts.

In summary, each of the three accounts described above offers an explanation for why non-nouns are understood around ten months, while nouns are understood around six months. The data-driven account suggests that this reflects nothing about the type of word, but rather reflects the statistics of these words’ use and appearance in infants’ lives. While a simple frequency-based version of this account is ruled out by our data, we find support for the idea that nouns are said more often in the presence of their referents than non-nouns. The social account suggests that infants’ developments in social cognition between 6-10 months are necessary for learning non-nouns, but perhaps not nouns; evidence from joint attention research is suggestive about the validity of this account. Ongoing research will help clarify the degree to which this account is correct. Finally, the conceptual account argues that there are properties specific to nouns that make them conceptually easier to learn than non-nouns, which can be conceived as a high- or low-level set of processes. Here too, ongoing research in our lab will help evaluate this hypothesis.

Growing Phonetic Precision

Another finding in this dissertation highlights an area in which infants’ early word comprehension is in a sense underspecified. As we saw in Chapter 4, 6-8 month old infants succeed in comprehending words for foods and body parts when they are labeled by their parent or by a new person, i.e. the experimenter, but fail to understand these words when the stressed vowel in them is changed. However, their performance at 6-8 months does not differ significantly from their performance when the words were correctly pronounced by either their mothers or by the experimenter. These results make it hard to make a strong claim one way or the other about infants’ early word-form representations. In contrast, by 12-14 months infants show a clear difference in
performance in the correct vs. mispronounced word comparison. In part, this is likely due to infants’ improving task performance around 13-14 months with correctly pronounced words, as we discuss in the next section. But another possibility is that early in word comprehension, infants’ word-forms are a bit more vague, more accepting of atypical instantiations for words they have learned.

This flexibility hypothesis has at least two possible implications. First, it may lend support to the idea that infants’ early word-forms are underspecified because they lack phonetic neighbors (e.g. ball & doll), as proposed by Charles-Luce & Luce (1990), though it is unclear whether more such neighbors are accrued between 6 and 12 months. Though the correlations computed in Chapter 4 suggest this is unlikely, one way to investigate this proposal further would be to conduct a corpus study looking at whether the number of lexical neighbors infants hear and/or understand grows between 6 and 12 months.

Second, the possibility of vague representations may lend support to the proposal that infants shift from a more acoustically driven to a more phonemically driven interpretation of speech sounds between 6 and 12 months (e.g. Jusczyk, 1993; Werker & Tees, 1984). Further research is needed to investigate this proposal. For instance, one might examine infants’ discrimination of the mispronunciations we used in Chapter 4, as well as their discrimination of different voices uttering our test words, to search for a shift in discrimination over development. A more acoustically driven process may lead to better discrimination of voices than of minimal pairs, given that at a gross level, the acoustics of different voices differ more than those between phonemes in a given voice. However, since indexical differences do not matter for meaning while phonemic ones do, it would be especially interesting to see if discrimination of the two kinds of differences changes over developmental time as word-learning rate and ability increases.

In summary, the present research leaves open the question of the nature of infants’ very early word form representations, but suggests that sometime between 6 and 12 months, the words in infants’ lexicons become phonetically well-specified.

**Improvement at 13-14 months**

Aside from the precocious yet staggered word comprehension found for nouns and non-nouns in this dissertation, another element of the data points to a striking parallel across word-types. Namely, infants’ comprehension abilities around 13 months show an increase for both nouns and non-nouns. That is, the degree to which infants are able to demonstrate their word meaning knowledge increases robustly, regardless of the word-type queried, and within our nouns, regardless of whether they appear in paired-picture or scene trials. What changes in cognitive and linguistic development lead to this increase in word comprehension?

Four possible explanations for infants’ increasing word comprehension abilities around 13 months, which may be explored in future research\(^\text{17}\), are linked to infants’

\[^{17}\text{Perhaps even my own.}\]
understanding of functional language elements, increasing speech production abilities, the onset of walking, and general cognitive and social development.

**Grammatical categories and word comprehension.**

Recent research suggests that infants and toddlers use function words, syntax, and sentence structure to direct their understanding of words more efficiently. For example, toddlers use Spanish article gender to guide their word comprehension, narrowing their attention to female-gendered nouns, e.g. “la pelota” (the ball) upon hearing “la” (Lew-Williams & Fernald, 2007). Similarly, 15 month olds (Jin & Fisher, in prep) and 2-year olds (Arunachalam & Waxman, 2010; Yuan & Fisher, 2009) use noun structure within a sentence to tell whether novel verbs involve two participants or one. Relatedly, work comparing infants’ abilities to link nouns with objects, and adjectives with properties, finds that the noun-object link emerges first at around 14 months, with robust adjective-property links emerging only after 18 months (Booth & Waxman, 2009). Thus, previous research suggests that not long after their first birthday, children start to harness structural information, especially from the syntactic structure that surrounds nouns.

Building on such previous demonstrations of budding grammatical knowledge and our word comprehension findings, one line of further research could explore the possibility that infants’ understanding of nouns as a grammatical category (i.e. one with word-order and function-word constraints) simplifies the task of content word comprehension. If this is so, it could be because the ratio of unknown to understood/expected elements in a sentence is tipped in infants’ favor. For example, while both 6- and 13-month-olds look at an apple when asked, “Do you see the apple?,” it could be that older infants show more robust performance in part because they’ve learned that “the” precedes forthcoming nouns, often objects, and that nouns often come at the ends of sentences. In contrast, the 6 month olds may essentially hear “bla bla bla bla apple,” and are responding to (rather than forming some sort of expectation about) the last bit of the sentence. One way to examine this possibility would be to manipulate sentence structure in a way that doesn’t impact meaning (e.g. by saying “See the do apple?”), to see if older infants perform less well with scrambled (and so less prediction-generating) word order. Thus, further research, perhaps along the lines just sketched, is needed to determine whether knowing about sentence structure helps word comprehension of nouns in particular (or conversely, if whatever causes increased noun comprehension then in turn facilitates knowledge about sentence structure).

However, distributional knowledge about function words is not likely the sole explanation behind the broader success of older infants that we find. During the non-noun study, only the barest syntax was used (e.g. Look! Dance, dance!), suggesting that older infants’ success in this study did not rely on a better understanding of grammatical elements or predictive listening in the sense described above (as in Lew-Williams & Fernald). Moreover, the group of words used in the non-nouns study was heterogeneous, ranging from social terms to performatives, adjectives, verbs, all of which have differing syntax (though there is no reason to think these words are in these word classes to infants). At the same time, many of these words occur commonly and often in isolation,
again suggesting the performance boost in this age group is not due to better use of grammatical cues.

Thus, a tight link between knowledge of the distribution of functional elements and increased word comprehension, if discovered, may only explain improved noun performance rather than our entire pattern of results. That said, it remains possible that older infants’ knowledge of functional elements is useful in a broader way, for language comprehension as a whole; these ideas await further research.

Word production and word comprehension.

A second possibility is that infants’ increasing comprehension is the result of the beginning of word production, which begins, on average, around infants’ first birthday. While it has long since been shown by researchers that comprehension precedes production (e.g., Benedict, 1979), the nature of this relationship is still underspecified. For instance, as Benedict documented (1979), the first 50 words understood and the first 50 words said were not the same words, nor were the relative proportions of different word classes in each type of vocabulary the same. Thus, given these differences between what words infants first understand, and what words they begin with when they begin to talk, there may be different propensities governing the contents of each type of vocabulary. At the same time, there has been a very well documented correlation between infants’ vocabulary size in both comprehension and production, as assessed by parental vocabulary checklist (Dale & Fenson, 1996; Fenson, et al., 1994). Thus, one avenue for future research is to explore this relationship further, to determine whether there is a boost in comprehension for a given infant observed when that infant begins to say words. Put otherwise, future studies could test the hypothesis that the word comprehension boost we find around 13-14 months in infants as a group occurs when infants gain insight that the words they themselves produce can be consistently linked with the referential world, and understood by others. ¹⁸

One could imagine different forms that such a production-comprehension link could take. For instance, the link between how well words are understood and whether infants are saying words could be binary, i.e. the comprehension boost could come from infants beginning to say any words at all, or a given number of words, regardless of which words they say. Or, infants might show a comprehension boost, but one that is specific to the words infants are saying, though the lack of correlations between parental vocabulary checklists and infants’ performance in word comprehension in our studies suggests this alternative of a production-comprehension link is somewhat unlikely (Bergelson & Swingley, 2012, 2013b)¹⁹. Yet another alternative is that the relation is general and graded, that is, the more words infants are saying, the better they understand words during comprehension in general. One way to test these alternatives is to test

---

¹⁸ This is speaking a bit loosely; I’m not suggesting infants need metacognitive awareness, but rather that whatever process leads to the onset of word production may be the same as the process that underlies more robust word comprehension.

¹⁹ Though this lack of correlation may have been driven by younger infants’ parents who generally thought their infants understood very few words, if any.
infants between 12-14 months on both comprehension measures and production measures, to investigate the nature of the links between them.

**Onset of walking.**

A third possibility is that the boost we find around 13 months is due to the onset of walking, which occurs for most infants between 11 and 15 months. Indeed, recent longitudinal research with 10-13 month olds (Walle & Campos, 2013) has documented a link between the onset of walking and infants’ receptive and productive vocabulary. These authors find that locomotor status (walking vs. crawling) predicted infants’ receptive and productive vocabulary above and beyond the effect of age, and that this relationship was non-linear. They also find that walking infants whose parents talk to them more during an observation have larger vocabularies than crawling infants whose parents talk to them more, and that walking infants of parents who moved less frequently during an observation had larger vocabularies, in addition to other effects of infant movement and distance-from-parent. These findings raise a fascinating possibility about the link between increased comprehension and locomotion, perhaps driven by infants’ newfound vantage point, the ability to carry objects that this grants, and their better opportunity to see parents’ direction of gaze (Franchak, Kretch, Soska, & Adolph, 2011; Soska, Adolph, & Johnson, 2010).

However, the results of Walle & Campos (2013) used parent vocabulary checklists (MCDI) rather than an in-lab behavioral test of infants’ word comprehension, and so are unable to speak directly to whether the onset of walking maps onto the improved comprehension around 13 months that we find. Given that parents of walkers attribute more maturity to them (as these authors suggest), one possibility is that this has an impact on how they fill out the MCDI (though, this study finds no differences between receptive and productive vocabulary in crawlers and walkers at 10.5 months, the onset of the study). Moreover, as discussed further below, overall vocabulary is not the same as word-comprehension proficiency, so an interesting future direction for this question would be to look for a link between onset of walking and word comprehension in an in-lab evaluation of both. Another way to get at this, as Walle and Campos suggest, is through more dense longitudinal video recordings of infants in their homes. One particularly effective way to examine this would be to compare infants’ visual input as crawlers and as walkers in the home through the use of head-mounted cameras (as Adolph and colleagues have done in the lab, e.g. Franchak, et al., 2011). This data could then be linked to infants’ language exposure and in-lab word comprehension as well.

**General cognitive/social development & word comprehension.**

A final possibility (perhaps better titled ‘the kitchen sink’) is that what leads to higher performance around 13 months is general cognitive and social development, rather than linguistic or motoric development per se. Indeed, across areas of developmental psychology, various social and cognitive skills such as gaze following and pointing (Brooks & Meltzoff, 2008), or memory and attention, *inter alia* (Rose, Feldman, & Jankowski, 2009) are predictive of infant vocabulary size. However, this work too generally uses a parental checklist as a measure of vocabulary size, and often has a large
delay between when a skill is tested and when the vocabulary is assessed (cf. Rose et al, 2009). In contrast, here the question of interest is about processing, i.e. what makes word comprehension more efficient rather than what predicts vocabulary size or growth. Word processing proficiency itself has been demonstrably tied to vocabulary growth in toddlers (Fernald & Marchman, 2012; Fernald, Perfors, & Marchman, 2006), but may independently relate to other factors, as can be tested by future studies. One likely set of candidate correlates for word comprehension includes the development of social attention, attention, memory, and representational competence.

In terms of social attention, as discussed in Chapter 3, it is only around 14 months that infants show true signs of ‘knowing together,’ i.e. truly joint attention, rather than parallel attention to the same object (Carpenter & Call, in press). It could be that an increase in this type of skill lets infants learn more from their day-to-day linguistic and environmental experiences, thus boosting their word comprehension.

In terms of attention allocation, rule-shifting tasks, such as those used with young infants to measure their ability to suppress a preferred or learned response (Kovacs & Mehler, 2009), could be used to help determine whether the top-down control of a learned behavior is used in word comprehension. This could be true in at least two ways. First, it could be that suppressing a prepotent response helps infants learn words because they are able to tune in to the linguistic, social, and contextual cues around them. Second, this ability could lead to better task performance, even if the knowledge of words is the same as in younger infants, because older infants could be better able to disengage from salient aspects of the image in the word comprehension task, and thus better able to show their knowledge. In light of our discussion of saliency issues during our experiments (see Chapter 4, Appendix A), this seems to be an important avenue of further research.

Both memory and representational competence have been found to be moderate predictors of receptive language, perhaps not unexpectedly (Rose et al, 2009). Memory, measured both in delayed and immediate recall, and in recognition tasks, has been found to predict concurrent and future receptive vocabulary. One can easily imagine that better memory overall could tie to better word-form and word-meaning representation, each of which are required for successful word comprehension. Representational competence, as measured in tasks involving cross-modal transfer, symbolic play, and object permanence, also predicts vocabulary measures. As suggested by Rose et al. (2009), “[these measures] all share with language the abstract mental representations of objects absent any immediate visual support” (p. 146). While this quotation suggests a link between representational abilities and words for objects, it may turn out that cross-modal transfer matters more directly for understanding speech as well (Yeung & Werker, 2013). Moreover, symbolic play and object permanence are each complicated constructs themselves: the former likely requires a high degree of abstraction and generalization, needed too for appropriately broad word representations and meaning extension; the

---

20 Studies using vocabulary checklists as a proxy of word comprehension have tested some aspects of attention, with mixed results (e.g. Colombo, 2004; Rose, et al., 2009) suggesting all the more that further research is needed to clarify the interaction of aspects of attention with word comprehension.
latter requires a broad understanding of objecthood and often intent as well. Thus, future work could unpack these complex abilities, to probe which aspects of them may support increased word comprehension, and why.

**Summary of 10-13 month improvement accounts.**

The cause of infants’ increase in word comprehension proficiency just after their first birthday, months after they first demonstrate word comprehension, is an important unanswered question for language development, and one whose likely answer involves interaction across levels of linguistic representation, and general cognitive development. The kinds of studies sketched above begin to address these issues by examining four ways this proficiency might be explained: a growing understanding of functional sentence elements, the onset of speech production, the onset of walking, and global social/cognitive abilities. The first step is to investigate the linkages between these areas, so that a careful causal account might emerge. Conducting such studies both across and within groups of infants, incorporating longitudinal designs, will lead to a rich data set that can then be probed through mediation analyses and data modeling. This can help determine which abilities predict comprehension measures. These results will provide a better understanding of how infants transition from understanding a few words to efficiently and proficiently understanding (and eventually speaking) their native language.

**Understanding the Conditions that Lead to Word Comprehension**

As I alluded to in the summary that opened this chapter, knowing when infants understand certain words allows us to build theories of their understanding, and design further experiments to test predictions of these various theories, but by and large, this work falls short of explaining *how* infants learn words. To get closer to understanding how infants comprehend words, it will be necessary in future work to bridge the gap between what we know about infants’ performance in the lab at different ages, and what their everyday experiences, both perceptual and linguistic, are like. We will then be in a better position to evaluate word-learning theories. This too, of course, is just a proxy of how infants learn, but getting a more multi-dimensional picture of infants’ daily lives and learning outcomes will lead to better fodder for testing mechanisms of word learning.

Indeed, a major challenge to research on word learning is the gap between research based on in-home recordings of natural interactions between infants and their parents, which provide highly heterogeneous and unbalanced data; and in-lab testing, which, while tightly controlled for many variables, is limited to querying a very small subset of infants’ knowledge due to limitations of experimental design (length of study, available methods, etc.) Moreover, current research points more and more to the importance of understanding the contribution of differences in daily experience on in-lab performance. One possible way to rectify this problem is to test the same infants in the lab that are recorded in the home. This type of approach combines the high-precision of questions answered by in-lab studies, and the more diffuse data-mining (dubbed “snowflake counting” and “butterfly collecting” by (Chomsky, 2010) garnered from longitudinal observational work. Put otherwise, such an approach would allow us to
combine information about infants’ exposure to certain words, sounds, and concepts with their internalization of this knowledge over time.

There are a few research questions in particular that would benefit greatly from a combination of at-home recordings and in-lab experiments. First, understanding particular infants’ experience with given common words (such as the nouns and non-nouns tested in this dissertation) may allow us to form predictions about their performance on individual items. That is, an infant who sees her mother peel and then feed her bananas for breakfast every morning might be expected to have learned the word banana better than the infant who only hears about or experiences bananas in the context of banana bread. While this example highlights a boring and obvious outcome, i.e. seeing an object more while hearing it labeled leads to higher levels of learning, it is likely the case that the link between exposure and learning is not this obvious. That is, one could imagine that what matters is not the quantity of exposure, but how clear, and clearly attended an exposure incident is. Perhaps the daily banana eater is playing with an iPad while mother is peeling and feeding her the banana. Perhaps the banana-bread eater had one very salient instance of passing back and forth a toy banana with a babysitter who labels it. The mapping between these kinds of experiences and the degree of infants’ learning is not yet clear. Through frequent at-home recordings, from the infants’ vantage point in addition to a more global one, followed by in-lab testing, we can gain a better understanding of what might underlie and predict a given child’s learning about a given concept.

A second benefit to this type of joint paradigm would be an ability to query infants’ exposure to, and subsequent comprehension of, words for novel items. By varying the similarity (phonological and/or semantic) between items introduced in the home and familiar items, we could gain invaluable insight into the process of lexicon expansion, word-form representation building, and semantic categorization. While a similar paradigm has been undertaken in purely visual or purely auditory tasks, confirming that the method is feasible (Jusczyk & Hohne, 1997; Scott & Monesson, 2009), this approach has never been used to query word-learning with novel (or familiar) words. Such an approach could potentially be far more informative than a brief teaching episode in the lab, where object name to referent mappings can be fleeting (Horst & Samuelson, 2008).

Another type of question answerable by a combination of at-home recordings and in-lab testing concerns visual feature learning, and more generally learning under complex visual conditions. While a good deal is known about how infants come to learn the speech-sounds of their native language, far less is known about how infants learn features like “animate” and “round,” or whether infants consider these features indexical for word meaning (e.g., whether roundness is part of the meaning of cookie; there has been great philosophical debate over this very issue (Fodor, 1998; Fodor, et al., 1980). Similarly, while much is known about how infants inspect and enlist their knowledge in sparse visual displays (Fernald, et al., 1998; Golinkoff, et al., 1987), we are only beginning to understand how infants pick out the relevant portions of the moving, complex world that they experience (Aslin, 2009; Frank, Tenenbaum, & Fernald, 2013), as we began to examine here with the scene trials and videos in Chapters 2 and 4. By
looking closely at the effects of visual saliency and linguistic input on infants’
allocated attention over development (facilitated by at-home recordings with both head-
mounted and fixed-location cameras), and the subsequent comprehension infants indicate
for attended items, we can begin to home in on the link between lower-level perceptual
processes and higher-order cognitive processes.

Thus the paradigms and research questions sketched here could be of great use in
better understanding the process of word-learning and the related degree of word
comprehension. Knowing about infants’ history with words, and linking this to how well
they show understanding of them, will contribute to the theories we can build about how
meaning, sound, and attention interact in early word-learning.

Word Learning Theories

Understanding how the present findings relate to theories of word learning is an
important challenge for future research. Below, I outline several such theories, and
provide some discussion of the implications for each given theory for early word
learning. While many mechanisms of word learning apply to older children’s learning,
i.e. mutual exclusivity, robust understanding of social intent, shape bias, etc., here I
constrain my discussion to two groups of theories that are meant to account for the word-
learning novice, who must, at least at first, learn through observation alone.

Cross-situational multiple hypothesis accounts.

One school of thought, exemplified by Smith and colleagues (Smith, 1995, 2000;
Smith & Yu, 2008; Yu & Smith, 2007) holds that word learning is built from simple
associative learning, the same kind of learning undergone by rats who link pressing a bar
with receiving pellets (Rescorla & Wagner, 1972). Under this type of model, infants learn
first through seeing consistent pairings between words and referents, and then use the
consistencies discovered there to develop word learning biases, e.g. the shape bias, which
in turn is used to learn further words. This kind of model is able, perhaps surprisingly, to
account for a wide array of findings, including 12-month-olds’ ability to learn multiple
referents under ambiguous exposure conditions (Smith & Yu, 2008), and young
children’s acquisition of the shape bias (Landau, Smith, & Jones, 1998).

There is some debate about how such a model would scale up to account for, e.g.,
the learning of words for imaginary, or non-present objects; general terms such as
‘animal;’ verbs and adjectives, etc., as well as concern that such a model should predict a
higher rate of mistakes, or mis-mappings between words and referents, than is found. For
instance, for the associationist infant with a forgetful parent, hearing ‘oh, let me go grab
your hat’ once the infant is already strapped into the stroller should, in principal, lead the
infant to learn ‘hat’ might mean ‘stroller’ or ‘door,’ since these items would regularly
appear in his field of view when hearing that word. And yet, cute anecdotes aside, these
kinds of errors are not often documented in the literature, and when they are, are often
compatible with an over- or under-extension of words on the part of the child, rather than
a mis-mapping (see Bloom, 2002 for further discussion).

Moreover, such an account would have trouble wholly explaining our findings.
While it’s possible that ‘bottle’ or ‘banana’ appear in infants’ experience with enough
regularity to be learned through rote associations, juice, milk, hand, and nose take on a wider range of visual forms that are less consistent, and surely have more nebulous boundaries or edges. Moreover, body parts offer a bit of a quandary for the purely associationist learner: some of them are very often present in the infants’ view, e.g. her own hands, and so should be likely possible referents for a great many words, while others can only be touched or seen on others, e.g. nose. And yet, these are some of the very first words infants learn. But even granting that the first words may be learned through associations between the word and its referent, associationist models would have to build on themselves and develop quite rapidly to explain how infants are able to learn ‘uh oh’ or ‘eat’ from repeated instances of these words in their environments.

Another account, exemplified by Xu and Tanenbaum (2007), suggests that infants learn words through rational inductive inferences, based on the statistics of their prior experiences with the words and their referents. These models do a good job accounting for toddler and adults’ rapid learning of new words after few instances, and, unlike associationist models, are able to account for learning of words within a taxonomic hierarchy. However, as Xu & Tanenbaum (2007) point out, early word learning appears far less efficient than later word learning, and as such may not depend on the same mechanisms, in their case Bayesian inference, or may reflect an immature implementation of these mechanisms. The robust learning we find in our older infants may reflect these infants’ more mature implementation of models along these lines.

One-trial learning.

A form of one-trial learning, championed by Trueswell, Gleitman and colleagues, proposes that rather than keeping track of multiple possible hypotheses for a word-referent link, which then get reweighted with incoming data, infants only posit one hypothesis at a time (Medina, et al., 2011; Trueswell, et al., 2013). This hypothesis is retained until and unless it is disproven. Thus if an infant sees a book and a ball for the first time, and hears ‘ball’, rather than an infant thinking that there’s a 50% chance ‘ball’ means ball and a 50% chance it means book, infants make a guess, e.g. that it means ‘book,’ and on the next learning instance, when there is no book present, discard the hypothesis in favor of a new one, perhaps ‘ball’ or perhaps ‘car,’ depending on the object of focus by the child and her parent.

This hypothesis, dubbed ‘propose-but-verify,’ accounts for cross-situational referent learning data of the kind examined by Yu and Smith (2007), as well as by other experiments, some using a more cluttered and naturalistic visual world (Medina, et al., 2011; Trueswell, et al., 2013). One idea raised by this hypothesis is that the number of potential referents is very large, and that associative, gradualist models may only be viable within a closed set. In contrast, one-trial learning, which has been found to be viable across domains and species (Gallistel, Fairhurst, & Balsam, 2004; Rock, 1957), may scale more appropriately to real-world learning situations. Indeed, as Trueswell et al. suggest (2013), much learning that looks gradual when only the end result of learning across a group is portrayed may in fact be more appropriately described as one-trial learning for each individual, with the point of learning differing across the population (see Gallistel, 2004 for further discussion).
Another benefit of the propose-but-verify hypothesis is that it predicts very slow early learning, much as we find from 6-13 months in Chapters 2-4. Then, when boosted by other linguistic cues, e.g. the onset of recruitment of syntactic and distributional information, or other factors we discussed above, this model is compatible with the fast increase in learning thereafter (Trueswell, et al., 2013).

The propose-but-verify hypothesis as a theory of early word learning is rather new, and as such still awaits empirical testing and further development. There are several issues it has yet to address. For instance, can this same mechanism be used to learn parts of speech other than nouns? That is, would it be a viable mechanism not only for, say, food and body-part words learned around 6 months but for the non-nouns in Chapter 3 as well? In principle there’s no reason why not, but just as in the multiple-hypothesis accounts sketched above, it has thus far only been tested as a theory of noun-learning. It’s not entirely clear how an infant even begins to posit other parts of speech, as discussed above: this is an area for further research.

Another area of development for this theory is that it does not yet specify how infants determine that a word-referent link has been correctly found, and that infants need not continue to test it further, or revise it. That is, as currently described, the authors suggest that it is possible that when encountering a situation that is not compatible with a confirmed hypothesis, that learners could actively search their memory for past hypotheses, and when the evidence supports it, learn multiple meanings for a given word form (Trueswell, et al., 2013, p. 154). While this takes the needed step of specifying how homophones are learned under this framework, it has that unwanted side effect of perhaps causing unlearning or uncertainty in the cases that mother says, ‘I’ll go get a ball’ with no ball present. Given that it is estimated that mothers’ utterances have to do with objects currently in the young infants’ view approximately 70% of the time (Harris, et al., 1983), there seems to be an very large set of data the child could use to discount her accurate learning.

A final problem, as the authors note, is that there is as yet no evidence to indicate that such a mechanism underlies word-learning by infants; all current evidence comes from adults (Trueswell, et al., 2013). This is a key area for further research.

**Word learning’s special status.**

There has been great debate in the field about whether word learning exemplifies a special kind of learning. The accounts just sketched all are undergirded by domain-general mechanisms (associative learning, Bayesian inference learning, and one-trial learning), though the word learning they establish may then give rise to language-specific patterns. For instance, the Bayesian model proposed by Xu & Tanenbaum offers a caveat that while the inference mechanisms are likely domain-general, “word learning may require certain language-specific principles or structures” (2007, p. 270).

Other accounts fiercely defend the ‘special’ nature of word-learning. These accounts have highlighted aspects of word-learning that distinguish it from learning other kinds of associations. This mainly concerns the ways in which newly learned words are generalized. Specialized generalization patterns are an early-appreciated aspect of words, in place to some degree by 6 months, when infants will learn a new object category if it is
linked to a word but not to a series of tones (Fulkerson & Waxman, 2007); a similar differentiation is found with 3 and 4 month olds (Ferry, Hespos, & Waxman, 2010). In toddlers, learning a fact (e.g. ‘my uncle gave me this’) and learning a label (e.g. ‘this is a fep’) manifests itself the same way in some regards, but facts and labels are differentially extended to novel category instances, and to expectations about community members’ shared knowledge (e.g. Diesendruck & Markson, 2001; Waxman & Booth, 2000).

However, it is not clear that the debate over the domain-specific nature of word-learning is fruitful for the current discussion, especially insofar as much of it has concerned learning by much older children than we investigate here, namely, 3-5 year olds. Moreover, it is not clear what is being argued over. Everyone seems to agree that there are some unique properties of words that are not true of other kinds of information, for instance, their patterns of generalization to other category members (Waxman & Booth, 2000). But it also seems to be the case that many of the mechanisms suggested for word learning are also used in other domains of cognition; true Fodorian modularity is no longer proposed as a serious cognitive architecture (Fodor, 1983; Waxman & Gelman, 2009).

In summary, there seem to be several mechanisms on the table that may prove viable for early word-learning either alone or in concert with other skills. Namely, cross-situational associative learning, multiple-hypothesis testing, or propose-but-verify single hypothesis testing could each be plausibly used to learn the most readily available words in infants’ linguistic and perceptual environments. While it is clear that before six months of age, infants are ready to preferentially link words with new object categories (Ferry et al, 2010), it is not clear whether, as currently formulated, the mechanisms described above would support the learning of foods and body parts, and soon thereafter, non-nouns like ‘uh-oh,’ and as such, each proposed hypothesis does not at present meet the desiderata for a theory of early word learning. Further research will be necessary to gain a better understanding of the plausibility of these theories to account for our data.

Cautionary Notes

In writing about young children’s early language production, Macnamara writes “…one jumps from a child’s language to his understanding at one’s own risk” (Macnamara, 1982, p. 161). I would add that the jump from the adult’s language to the child’s understanding is equally perilous. There are several things to keep in mind about what the studies documented in this dissertation have and haven’t shown, and what we know about whether the various mechanisms I’ve discussed are necessary or simply useful for word-learning, when they’re available.

In the studies described in Chapters 2-4, we have shown that young infants have some word comprehension knowledge, which becomes more robust around 13-14 months of age. However, from these data alone it is not clear how adult-like infants’ word knowledge is, especially in terms of categorization, word-class, and conceptual knowledge. All we can say with confidence is that using these stimuli, groups of infants of given ages looked more at the correct image or video when its name was said (and moreover, only when it was properly pronounced). However, it is not clear from our tasks whether infants would identify these words in all and only the circumstances that adults
would. That is, infants may not classify a dog’s foot as ‘foot,’ or a purple apple as ‘apple,’ or a social faux pas as ‘uh-oh,’ until later in development, though in these cases adults too would perform less well than when these words are depicted in canonical form (Rosch, et al., 1976). Nor do we have evidence that there is any sort of syntactic word-class knowledge in infants’ word representations at these young ages. The competing lexical items in our paired-picture and paired-video trials are purposely quite different, and leave open the possibility that infants know something about the meanings of these words, but lack a fully specified lexical entry, however defined. Thus, our results should be interpreted and extended with caution, as modest evidence of word comprehension in infancy.

Another issue to keep in mind in our discussion of how children learn words is that several of the mechanisms discussed here, while certainly helpful for word-learning, are not strictly speaking necessary for word learning. That is, given the opportunity, children use gaze-cues to meaning (Baldwin, 1993), but in the absence of such cues, the blind learn language on more or less the same timeline (Landau & Gleitman, 1985). Similarly, while infants prefer and perhaps learn better from infant-directed speech (Fernald & Kuhl, 1987), infants whose parents never speak to them in this register, or even much at all, as is the case in some cultures, nevertheless learn language too (Schieffelin & Ochs, 1998); though, at the same time, there is a cost of hearing less input as well (Fernald, Marchman, & Weisleder, 2012; Hart & Risley, 1995). As a final case, though infants early-achieved successes in discriminating speech sounds, word classes, patterns, etc. as neonates has been well-documented, and is often seen as an important pre-requisite for language learning, they are not true pre-requisites for learning language. Deaf children, who are not privy to the same speech input, when exposed to sign or a new language at an early enough age (i.e. within the first five years of their life, give or take), nevertheless become completely fluent in sign language (Johnson & Newport, 1989; Newport, 1990).

Without going as far as these perhaps more rare and exceptional circumstances, there is further evidence that the timelines mentioned throughout this dissertation are not set in stone: infants who are exposed to two languages at once, or who are internationally adopted or move in early or middle childhood and have to start over with a brand new language, nevertheless sort out the relevant linguistic system(s), and often gain levels of fluency indiscriminable from adults’ (Pallier, et al., 2003; Snedeker, Geren, & Shafto, 2007).

These examples are not meant to discount the massive evidence concerning the use of visual cues or spoken speech, and their relations to word-learning, but just to remind the reader (and the author) that the mechanisms we propose and attempt to verify through our research are rarely as steadfast as we often assume.

Conclusions

The findings in this dissertation, namely, that infants understand common nouns for foods and body parts from 6 months onward, that they understand non-nouns a few months thereafter, open up myriad avenues for further research. The precocious word comprehension we find between 6 and 13 months in chapters 2-4 raises the bar for further
studies to uncover what the true necessary and sufficient prerequisites are for learning words. Moreover, the learning mechanisms uncovered in future work must give rise to infants’ learning of foods and body part words by around 6 months, and, perhaps in a harder task, non-nouns not long thereafter. Finally, we still only have hints as to what might give rise to the highly robust word learning that we find around 13-14 months, across word types and visual presentations.

By building and testing theories, and collecting more data, especially data that incorporates observational and experimental aspects, we can begin to answer some of these questions. At the same time, these findings can and should be extended to deepen our breadth of understanding of how word comprehension progresses, for instance, in the case of bilingual learners, second-language learners, and learners in various clinical populations. These findings, in turn, can be used to develop better diagnostic tools for understanding and identifying language deficits in pediatric populations, and for developing relevant interventions.
BIBLIOGRAPHY

Chapter 1 References:


Chapter 2 References:
These references retain the bibliography format of the original publication in PNAS


Chapter 3 References:


Chapter 4 References:


Chapter 5 References:


