



ELSEVIER

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Modeling the role of distributional information in children's use of phonemic contrasts

Erik D. Thiessen ^{a,*}, Philip I. Pavlik Jr. ^b^a Carnegie Mellon University, United States^b University of Memphis, United States

ARTICLE INFO

Article history:

Received 1 December 2014

revision received 13 January 2016

Keywords:

Statistical learning

Word learning

Exemplar memory model

Memory

Abstraction

Language

ABSTRACT

Between the first and the second year of life, children improve in their ability to use phonemic contrasts when learning label–object pairings. This improvement may be related to children's experience with the distribution of phonemes across lexical forms. Because phonemes typically occur in different lexical frames (e.g., /d/ and /t/ in “doggy” and “teddy” rather than “doggy” and “toggy”), familiarity with words makes similar phonemes more distinct through acquired distinctiveness. In a series of simulations, we demonstrate that English input has the distributional characteristics necessary to facilitate use of phonemic contrasts as a function of increasing familiarity with the lexicon. Further, these simulations support a novel prediction: that less common phonemes should take longer to be used productively. We tested this prediction with children between 18 and 25 months, and found that the relatively infrequent /s/ and /z/ contrast takes longer to emerge than frequent contrasts such as /b-/d/ or /d-/t/.

© 2016 Elsevier Inc. All rights reserved.

By the first birthday, the average monolingual infant is already familiar with the meaning of nearly 100 words (Fenson et al., 2002). But despite infants' success in discovering word–object associations at this age, they appear to have difficulty taking advantage of phonemic distinctions when making connections between word forms and meaning. One striking example of this is infants' failure to respond differentially to lexical forms that differ by only a single phoneme (i.e., *minimal pairs*) in label–object association tasks. For example, 14-month-old infants who have learned (via habituation) that a novel object is associated with the label /da/ respond equivalently when that object is labeled /ta/ as when it is labeled /da/. This failure to differentiate between the labels is not limited to /da/ and /ta/, and has been replicated with a variety of phonemic contrasts in word–object association tasks (e.g., Pater, Stager,

& Werker, 2004; Thiessen & Yee, 2010). Infants' failure to respond to these phonemic differences is not due to an inability to hear them, but is instead specific to settings where they are asked to use these contrasts to differentiate word meanings (e.g., Stager & Werker, 1997; Swingley & Aslin, 2000; Thiessen, 2007; Yoshida, Fennell, Swingley, & Werker, 2009). Instead of a perceptual failure, infants' failures to use phonemic differences in label–object association tasks appear to result from a difficulty in *making use* of some phonemic distinctions that they can perceive and encode (e.g., Shvachkin, 1973; Thiessen, 2011).

That is, though infants perceive and encode the phonemic distinctions that are relevant to their native language when they are building a lexicon, they initially fail to treat these phonemic distinctions as signifying the distinction between tokens of different lexical categories (Swingley & Aslin, 2007). Over the course of development, this difficulty is substantially alleviated. By 18 months, infants use phonemic distinctions (at least, the word-initial stop consonant contrasts typically used in laboratory tasks)

* Corresponding author at: Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, United States.

E-mail address: thiessen@andrew.cmu.edu (E.D. Thiessen).

contrastively in word–object association tasks (e.g., Thiessen, 2007; Thiessen & Yee, 2010). Three potential explanations have been forwarded for this developmental change. The first is a capacity account. On this account, young infants' failures are due to the fact that the task is demanding – it requires infants to encode the label, the object, and the link between them. Because young infants have less capacity than older infants and adults, they cannot encode all of this information simultaneously, and subsequently fail to detect subtle changes in the identity of the label. Older infants succeed because they have more capacity (Werker, Fennell, Corcoran, & Stager, 2002). The second is a social inference account. From this perspective, the reason that infants fail is that the label–object tasks used in the lab lack sufficient social support for infants to react to the labels as though they are meaningful. As such, infants ignore distinctions that they would treat as informative in a “real” or “linguistic” label–object association setting (Fennell & Waxman, 2010). While it is clear that social factors influence infants' word learning, this perspective does not make strong claims about what differentiates younger infants (who fail to use phonemic distinctions in laboratory tasks) from older infants (who succeed).

A final account suggests that infants' success and failure in using phonemic distinctions in laboratory tasks is due to their experience with the distributions of phonemes in their native language. In particular, Thiessen and colleagues (Thiessen, 2007, 2011; Thiessen & Pavlik, 2013; Thiessen & Yee, 2010) have proposed that experience with lexical forms helps to differentiate phonemic contrasts. On this account, infants are encoding a great deal of acoustic information associated with lexical forms. In addition to phonemic identity, for example, infants also encode indexical information about the speaker of a word (e.g., Houston & Jusczyk, 2003; Singh, 2008; Werker & Curtin, 2005). Initially, it may not be apparent which aspects of the acoustic variability in the encoded word forms are relevant for differentiating among different tokens of spoken words. Experience with lexical forms is informative to the extent that these forms help infants resolve the ambiguity inherent in the perceptual input and lead infants to weight phonemic distinctions more heavily (e.g., Swingley, 2009). This resolution occurs because experience with lexical forms is not random. Instead, infants are especially likely to encounter phonemes in distinct lexical contexts (such as /d/ and /t/ in *doggy* and *teddy*) as they develop a lexicon. Compared with the adult lexicon, children's lexicons contain fewer words where phonemes occur in identical contexts (i.e., fewer minimal pairs, such as /d/ and /t/ in *dip* and *tip*). For example, Swingley and Aslin (2007) found that over two thirds of the words in the vocabularies of 18-month-old Dutch learning infants had no minimal pair neighbors. Similarly, there are no single-feature minimal pairs in the first 50 words that children are most likely to comprehend (Caselli et al., 1995).

The experience of phonemic contrasts in distinct lexical contexts may serve to differentiate the contrasts due to a process known as *acquired distinctiveness*. Two similar stimuli, when paired with distinct outcomes, become more differentiable (e.g., Hall, 1991). That is, if an organism has

difficulty differentiating between stimuli *A* and *B* (i.e., two similar phonemes), they can be repeatedly paired with two more easily distinguished outcomes *X* and *Y* (e.g., *X* might be reward and *Y* punishment), such that the organism consistently experiences *AX* and *BY* pairings. Over time, these pairings reinforce the original (difficult) distinction between *A* and *B* and the distinction becomes more robust. In the child's developing lexicon, phonemes that are initially difficult to distinguish become more differentiable as they are paired with distinct lexical contexts. Because children know so few minimal pair words, they are unlikely to experience two phonemes in identical contexts (like /d/ and /t/ in *dip* and *tip*). Instead, they experience phonemes primarily in distinct contexts, and this experience with lexical could potentially help to differentiate similar phonemes.

To test this hypothesis, Thiessen (2007) conducted a laboratory training procedure intended to facilitate children's use of a phonemic contrast in a word–object association task. In that procedure, 15-month-olds who typically fail to respond differentially to the /d/–/t/ distinction in a word–object association task were exposed to the contrast in distinct lexical contexts, /dabo/ and /tagu/. After exposure to these labels, infants succeeded in responding differentially to the labels /da/ and /ta/ (for a replication, see Thiessen, 2011). This result is not simply due to increased familiarity with /d/ and /t/. When infants were exposed to these consonants for the same amount of time in identical lexical contexts (such as /dagu/ and /tagu/), they showed no benefit from the exposure and continued to respond to /d/ and /t/ as though they were interchangeable. These results are consistent with the hypothesis that exposure to phonemes in distinct lexical contexts helps to make similar phonemes more differentiable. More generally, this suggests that the distribution of phonemes in lexical forms, and infants' increasing familiarity with those lexical forms, plays an important role in infants' improving ability to make use of phonemic contrasts in word–object association tasks (e.g., Stager & Werker, 1997; Werker et al., 2002).

Of the three accounts discussed here, the distributional experience account makes a unique prediction: it suggests that children's ability to make use of particular phonemic contrasts in a label–object association task should be predictable from the frequency and distribution with which children have experienced those specific contrasts. That is, when two phonemes are similar to each other – for example, phonemes that differ by only a single phonetic feature such as voicing, which we will refer to as “minimal pair” phonemes – children need to experience the phoneme in distinct contexts for the members of the phoneme to become distinct enough to respond to them differentially in a label–object association task. This will take different amounts of time for different phonemes, as a function of the frequency with which children experience them in their language, such that children should succeed with more frequent contrasts (such as /d/ and /t/) before they succeed with less frequent contrast (such as /s/ and /z/). By contrast, the capacity account (Werker et al., 2002) suggests that children should succeed with all phonemes once they have enough capacity to encode the

information in the task. While the social support account does not make explicit developmental predictions, to the extent that its causal claims relate to children's understanding of the experimental setting, it should similarly predict that children succeed on most or all phonemes simultaneously: once children understand that the experimental setting is a communicative setting, they should treat the labels (whatever their phonemic content) as informative.

We will explore the predictions of the distributional learning account in two interrelated ways. First, we will conduct a simulation using iMinerva, a computational approach that is capable of simulating this kind of learning (Thiessen & Pavlik, 2013). The simulation will allow us to verify, first, that the distributional properties of the plausible input to children are such that phonemic contrasts will become more distinctive – and thus more likely to be responded to differentially – over the course of experience with lexical forms in the input. Second, the simulation will lend quantitative precision to the intuitive prediction that some contrasts should emerge later than others. Using the stimulation as a guide, we will also run a series of behavioral experiments, using a common variant of the label-object association task (e.g., Stager & Werker, 1997; Thiessen, 2007) to assess the prediction that less frequent contrasts emerge later than more frequent contrasts. While this prediction is intuitive, and has been explored in domains such as speech perception (e.g., Abbs & Minifie, 1969; Anderson, Morgan, & White, 2003) and production (Nissen & Fox, 2005), to our knowledge this is the first investigation of the role of phoneme frequency in children's differentiation of phonemic contrasts in word-object association tasks.

Introduction to iMinerva

In previous work, we have found that the iMinerva model is capable of simulating the facilitative effects of distinct artificial lexical contexts (like *dabo* and *tagu*) on phonemic contrasts in laboratory setting (Thiessen & Pavlik, 2013). iMinerva is an exemplar memory model that relies on four interrelated processes: similarity-based activation, integration of current experience with prior exemplars, decay of previously experienced exemplars over time, and abstraction. Like all exemplar memory models, iMinerva stores prior experience in the form of discrete exemplars, which are coded as n -dimensional vectors with positive and negative feature values. Each feature describes some characteristic of the stimuli, and the valence of the feature describes whether the characteristic is present (positive values) or contradicted (negative values). For example, a vector description of a square would have a positive value for the feature 4-sided, and a negative value for the feature 5-sided. The magnitude of the feature describes the degree of certainty that the feature is present or contradicted. Due to the effect of decay, magnitude has a tendency to converge toward zero.

When a new exemplar is presented to iMinerva, previously stored exemplars are activated as a function of their similarity. Similarity between exemplars is computed on a

feature-by-feature basis, considering both the magnitude and valence of each feature (for a complete description, see Thiessen & Pavlik, 2013). The most similar of these prior exemplars then “engages” with the current exemplar to create an interpretation (a new third vector that is stored in memory) of the current experience in light of prior experience. Interpretations are created through an additive merging of the current exemplar with the most similar prior exemplar. If the current exemplar has features that are consistent with the exemplar, then the engagement increases the magnitude of these features in the resulting interpretation. If the two vectors have features with opposing valences, then the interpretation resulting from their engagement will have less extreme feature values than the original exemplars.

Interpretations are stored in the same manner as exemplars, and the creation and storage of these interpretations yields sensitivity to the central tendency of exemplars in the input. For example, consider what would happen if iMinerva were exposed to a series of two-feature vectors, where the first feature of all of the vectors is 1, while the second feature alternates between 1 and -1 . Across a series of engagements, the first feature will be reinforced and increase in magnitude. In contrast, the magnitude of the second feature will decrease, because the engagement between current and prior vectors will often have inconsistent valences. Over time this inconsistency, coupled with the effect of decay, will lead the value of the second feature of novel interpretations to converge toward zero. Features that fall toward zero will not be retained forever in iMinerva's representation. iMinerva transforms features to null values when an interpretation contains features whose values fall below some fraction (determined by a parameter) of the average absolute feature strength for that interpretation. Features that are nullified in this manner are no longer used to compute similarity ratings, which facilitates generalization to novel stimuli that vary on “uninformative” features (e.g., McClelland & Plaut, 1999). The abstraction process simulates the fact that experience often results in a decrease in similarity to certain features of the input (e.g., Werker & Tees, 1984). As this example illustrates, iMinerva continually refines its interpretations in a way that is consistent with the central tendency of the input, and emphasizing those features that are consistent and abstracting away features that are inconsistent.

The four processes invoked by iMinerva – similarity based activation, engagement, decay, and abstraction – allow iMinerva to discover distributional regularities in the input. One of these distributional regularities is the lexical context in which phonemes occur. iMinerva responds to /d/ and /t/ as more similar when they consistently occur in the same context than when they consistently occur in different contexts (Thiessen & Pavlik, 2013). However, that simulation was based entirely on laboratory materials. In the set of simulations described below, we extend this principle to simulations based on the most common words in a developing child's English lexicon, as estimated by the MacArthur-Bates Communicative Development Inventory (MBCDI, Fenson et al., 2002). The characteristics of the lexicon will lead to behavioral predictions that we will test in a population of 15-month-old children.

Simulation

Thiessen (2007) found that exposure to /d/ and /t/ in distinct lexical contexts (/dabo/ and /tagu/) facilitated use of the /da/-/ta/ contrast in a word-object association task, while exposure to /d/ and /t/ in identical lexical contexts (/dagu/ and /tagu/) did not. Thiessen and Pavlik (2013) have demonstrated that iMinerva is able to simulate this effect. When the model is exposed to /dabo/ and /tagu/, a test probe of /da/ activates a different interpretation (one that incorporates traces of the second syllable /bo/) than a test probe of /ta/. However, if the model is exposed to the syllables /da/ and /ta/ in identical contexts (/dagu/ and /tagu/), a test probe of either syllable activates the same interpretation. That is, the model's response to /da/ and /ta/ is more similar after exposure to /dagu/ and /tagu/ than is its response to the same syllables after exposure to /dabo/ and /tagu/. This result is consistent with the performance of infants after exposure to the same sets of lexical contexts (Thiessen, 2011).

However, prior iMinerva simulations of distributional learning were focused on the input children receive in the experimental setting. In this set of simulations, our goal is to investigate two interrelated predictions. The first is that the characteristics of the lexicon support distributional learning, such that the words to which children are exposed make minimal pair phonemes more differentiable over the course of experience. Prior investigations of the developing lexicon support this prediction. Phonemes become more differentiable when they occur in distinct lexical contexts (e.g., /d/ and /t/ in “doggy” and “telephone”) rather than identical lexical contexts (e.g., /d/ and /t/ in “door” and “tore”). Children, however, know very few of these kinds of identical context word pairs (e.g., Caselli et al., 1995; Charles-Luce & Luce, 1990, 1995; Coady & Aslin, 2003; Swingley & Aslin, 2007). Instead, most of the words with which children are familiar provide evidence of phonemes occurring in distinct contexts. In this simulation, we will investigate the distributional properties of common lexical forms by looking to see whether they provide evidence that would make the representation of phonemes more distinct. That is, phonemes that are initially confusable (similar in representational space) should gradually become more differentiable as children accrue more experience with their distribution in distinct lexical contexts.

Because we propose differentiation occurs gradually, a more frequent pair of phonemes should be more differentiable than a less frequent pair. This simulation explores this prediction by comparing the degree of differentiation for the /d/-/t/ voicing contrast and the /s/-/z/ voicing contrast. The /d/-/t/ contrast is more frequent than the /s/-/z/ contrast in word-initial position (we focused solely on occurrence in word-initial position because prior experimental work suggests that children do not readily generalize distributional information across word position; Thiessen & Yee, 2010). The MacArthur CDI lists 48 words that start with either /d/ or /t/, and each phoneme occurs in word-initial position for over 20 words on the CDI. Words beginning with /s/ are also quite common (there

are 34), but children are exposed to very few /z/-initial words. The only two such words on the MBCDI are “zoo” and “zipper.” Because of this distribution, an infant has ample opportunity to learn that /d/ and /t/ occur in different lexical contexts, but less experience with the fact that /s/ and /z/ occur in different contexts due to the fact that infants have comparatively little information about the contexts in which /z/ occurs.

Critically, though, the effect of phoneme frequency should not be estimated solely on the basis of the number of words in which a phoneme occurs. The token frequency of those words also plays an important role. In this simulation, we will vary how often each individual lexical form occurs, to simulate the fact that some words occur more frequently than other words. Most word frequency estimates have been derived from adult corpora, which necessarily have some systematic differences from corpora of utterances directed to infants and young children. In this simulation, we used the CHILDES database to calculate word frequencies of the items that children are likely to know, according to the MBCDI. The token frequencies of these items were then used to create an input to our simulated learners in which the proportion of token frequencies in the simulated input matched the proportion of token frequencies estimated from CHILDES.

Method

Stimuli

The model was presented with 800 tokens of 321 possible mono- and bisyllabic words present on the MBCDI. These words were translated into feature vectors using a similar coding scheme to one we have used previously (Thiessen & Pavlik, 2013). Each syllable was represented by 74 features, 14 each for the 2 possible initial consonants, 18 features for the vowel, and 28 more features for the 2 final consonant. In the case of our set of one and 2 syllable words (the MBCDI words beginning with /d/ and /t/), we represented each word with a vector of 148 values, where 0 indicates the feature is not present, 1 indicates a positive valence of the feature (e.g., voiced) and -1 represents a negative valence of the feature (e.g., unvoiced). Features of the input always range between -1 and 1, while features in the model's memory interpretations may take any real values since their strength measures the salience of the feature in the model's interpretation. Additionally, when features have low strength in the representation they are set to null to simulate abstraction processes; however this mechanism is not important to the results in this paper (because we are not assessing abstraction) and we set the abstraction parameter at the typical default from Thiessen and Pavlik (2013), as shown in Table A2 in the appendix.

Procedure

We ran 250 simulations of exposure to our corpus; each simulation models an individual learner. The corpora for each simulation were not identical. Instead, words were selected according to their probability of being directed at a child as computed from their relative frequency in

the CHILDES database. As such, while the corpora for each simulation are generally similar (e.g., more frequent words are more likely to occur), individual simulations experienced individualized input. We probed each simulation's knowledge after it had received 200, 400, 600, and 800 tokens. Note that these numerical values are not intended to be indicative of the actual number of times that a child would need to hear a particular word; instead, they are selected to provide an appropriate number of exposures for the model to learn. For each individual simulation, we analyzed how that simulated infant differentiated the four phonemes of interest (*/d/*, */t/*, */s/*, and */z/*) at these four timepoints. During testing, we found our decay parameter was excessive for long simulations (because the input took much longer to unspool over the vastly increased number of exemplars) and decay was much reduced compared to Thiessen and Pavlik (2013). For precise specification of all parameters, see Table A2.

Our primary prediction was that discrimination of the */d/-/t/* contrast would emerge sooner than discrimination of the */s/-/z/* contrast. To model discrimination we needed to develop a novel way to compare the echo resulting from a probe with the echoes from other possible probes, while also accounting for the effect of strength (i.e., frequency) in a plausible manner. This function is described in detail in the appendix; what follows is a general overview. The first step in the function is to calculate the similarity of the specific probe with echoes from all 4 phonemes. This value is then normalized by dividing the resulting similarity values by the highest similarity value (which in all cases is an identical phoneme; i.e., out of the 4 possible phonemes, */d/* is always most similar to */d/*). This means that the maximum normalized similarity is 1, and gives rise to similarity scores ranging between 0 (no featural overlap) and 1 (perfect overlap).

Then, this value is raised to the exponent of the absolute value of the vector (overall memory strength), which compresses the similarities according to their strength. This reflects a memory resonance process (MacWhinney, 2004), where stronger comparators have a high power, causing multiple resonant comparisons, each of which reduces the perceived similarity with the dissimilar probe (because the correct probe has a similarity of 1, this process only depresses incorrect matches as a function of their strength). Finally, we also multiplied this compressed similarity by the overall memory strength to capture the other part of the effect, the increased “volume” or salience of the stronger memory. This occurs because the stronger competitors compete better regardless of their similarity, in much the same way that more frequent words are recognized more easily.

This equation works well because the general positive linear effect of strength early in learning is overwhelmed by the discriminative exponential resonance of strong memories. We will see this equation produces a pattern where discrimination is largely controlled by the strength (i.e., frequency) of a category of exemplars early on, but then as strength accumulates pattern match becomes more important. We will measure the discrimination for each probe by computing this recall strength discrimination value for each of the 4 echoes. We do this for each of the 4 possible phoneme echoes (which we consider to

be the long-term memory trace for each phoneme) to see which is most strongly active (or discriminated) given the presence of each probe.

Results and discussion

Fig. 1 shows a four panel graph where the four panels represent */d/* and */t/* probes in the top left and right respectively, and */s/* and */z/* probes in the bottom left and right respectively. The values were plotted for simulations of 250 children for conditions of 200, 400, 600 or 800 lexical tokens. The figures graph the simulations' recall strength for each of the phonemes when probed with the test phoneme. The recall discrimination strength represents the propensity to identify that representation as fitting in the */d/*, */t/*, */s/* or */z/* categories. It is important to note that our use of 200, 400, 600 and 800 example trials greatly underrepresents how many words children actually see. For this reason we again emphasize the purpose of this result is to show the principles behind our explanation of the human data, not to provide a tight quantitative fit.

Fig. 1 illustrates two major conclusions. The first is that the degree of perceptual difference in the phonetic input matters to iMinerva. Early in learning, the model often confuses phonemes that are differentiated by only a single phonetic feature (i.e., */d/* and */t/* are only differentiated by voicing, as are */s/* and */z/*). By contrast, the model never confuses phonemes that are differentiated by multiple features (e.g., */s/* and */d/* differ not only on voicing, but also on manner and place of articulation). While the role of the degree of perceptual distinctiveness on children's use of phonemic contrasts in the Switch task has not been exhaustively investigated, these simulations are consistent with evidence that more similar phonemes present more difficulty for children in the Switch task (White & Morgan, 2008).

The second conclusion is that, for confusable contrasts (here, */d/-/t/*, and */s/-/z/*), discrimination improves with exposure. The rate of this improvement is related to the frequency with which the model experienced the phonemes in the input corpus. The frequent phonemes in the input (*/d/*, */t/* and */s/*) have positive discrimination early, which improves with time. While minimal pair phoneme strength also grows with learning, the difference between the correct phoneme (e.g., */d/*) and its minimal pair (e.g., */t/*) increases over experience, and the degree to which a phoneme activates its minimal pair begins to plateau around 800 trials, while the degree to which the phoneme activates itself continues to increase across the entire corpus.

Finally, as predicted, the low frequency comparator */z/*, at 200 trials is dominated by the strength of the */s/* memories, which are recalled more strongly. We can see that this difference flattens out across the 400 and 600 trial tests, as the simulations gained stronger representations, and that finally we see at 800 trials that */z/* is the dominant chunk recalled for */z/* probes.

Exploring the parameter space

While developing our discrimination function, we discovered that the model's discrimination between

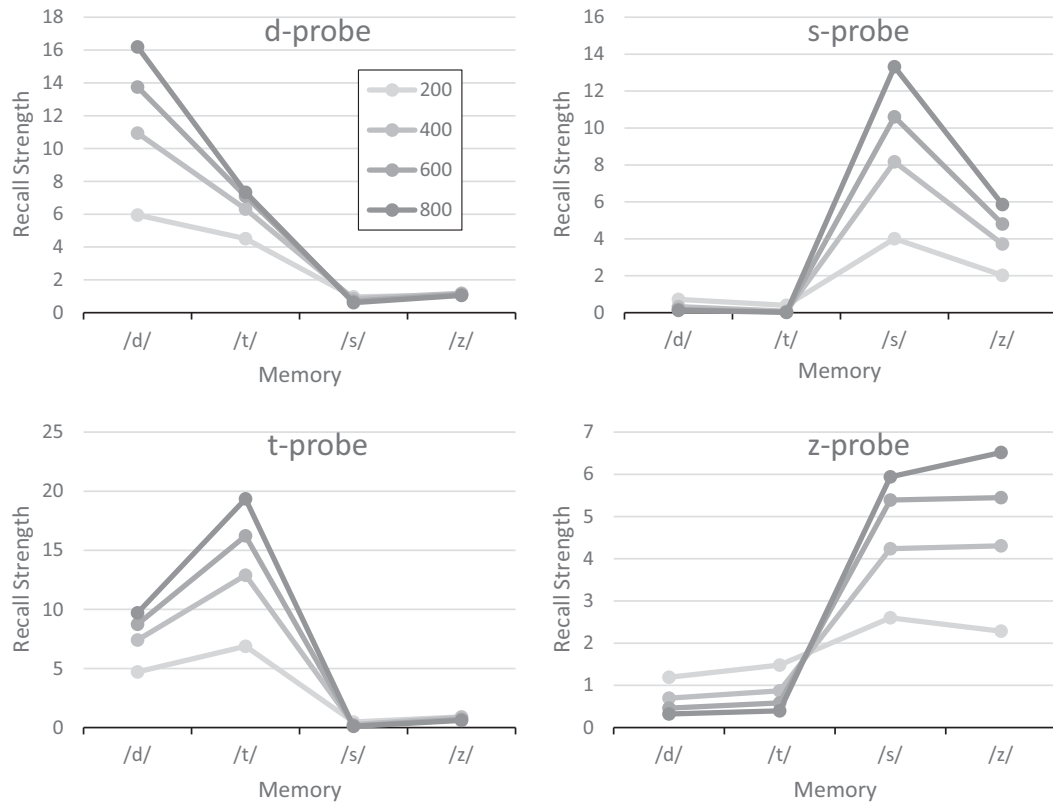


Fig. 1. Panels show the activation of memories for the 4 phonemes in response to each of the 4 probes for simulations of 200, 400, 600 or 800 trials.

phonemic probes was sensitive to the initial settings of the model in addition to the feature vectors the model experiences. Assuming a similar principle in humans, this suggests that the initial capabilities of a person, due to individual differences, might lead to a significant difference in model behavior. This implies that the model parameter values do influence the result and we might expect individual difference in parameters may account for model behavior difference. As an example, consider the decay parameter. The model produces excellent learning if decay is set low (in the extreme, if set to 0, the model would forget nothing), but poor learning if the decay parameter is set high (in the extreme, if set to 1, the model would instantly forget every lexical item it had ever saw, and never be able to learn any discriminations). This is intuitively plausible, in that it should be more difficult to learn to discriminate a contrast that one cannot remember experiencing.

As this discussion indicates, iMinerva's parameter settings have an influence on the rate at which (and the degree to which) the model learns. This raises the possibility that the results of the prior simulation are idiosyncratic to the particular parameter settings we used, rather than a general feature of the learning processes simulated in iMinerva. To assess this possibility, we performed a lengthy second set of simulations, using the same stimuli and amounts of practice as in the prior simulations, with the goal of determining how "widespread" the key result

of the simulation (faster discrimination of more frequent minimal pair phonemes) is across the parameter space.

We chose a parameter space to search where each of the 7 parameters can take one of 3 values, and we simulated 30 subjects at each setting of the 7 parameters. Table A2 shows the searched values for each parameter. Table A2 shows the searched values for each parameter. We searched the entire 2187 models indicated by the three to the seventh power possible model combinations of parameters settings. The results were revealing both for showing the importance of some parameters and the insensitivity of the results for other parameters. First, we observed that at all settings /s/ was always correctly identified with the /s/ iMinerva interpretation. Within our parameter ranges then, there were no cases where different parameters could result in the perverse result of /s/ being identified as a /z/. In contrast, for all settings we tried, we saw that only 64% (1391) of the settings resulted in /z/ correctly identified with the /z/ iMinerva interpretation given 800 trials. The parameter settings for these models are shown in Table 1.

For these 64% of the models, 68% of the models showed /z/ interpretation for /z/ presentation was greater already by trial 200, which indicates 32% of the 64% showed /z/ interpretation that tended to fail at 200 trials, but tended to succeed at 800 trials. Digging deeper into these 444 models (shown in Table 2) we can see some slight trends toward crossover when decay is less (e.g., .999) and when learning rate is less (e.g., .05). Abstraction parameter and

Table 1

Results for all simulation parameter settings that end with a /z/ recognition.

Model parameter	Low	Medium	High
λ	432	473	486
ρ	447	459	485
δ	465	463	463
Engagement threshold	463	466	462
Engagement threshold noise	462	465	464
Retrieval threshold	729	486	176
Retrieval threshold noise	243	486	662

Table 2

Results for simulation parameter settings that begin with a /z/ recognition deficit, and end with a /z/ recognized.

Model parameter	Low	Medium	High
λ	156	161	127
ρ	167	145	132
δ	151	148	145
Engagement threshold	142	155	148
Engagement threshold noise	145	151	148
Retrieval threshold	0	268	176
Retrieval threshold noise	0	203	241

the engagement threshold, and its noise parameter show no indications of effects. In contrast, the retrieval threshold was highly relevant, since none of the models showed the crossover to correct /z/ recognition when retrieval threshold was low. Rather, a mid-range value for the retrieval threshold tended to correspond with this crossover effect. We also saw that the retrieval threshold noise was an important factor, with this crossover to the correct understanding occurring only when noise was higher.

It is also very relevant to consider the 36% of the models where /z/ was not learned (model fails). These models are shown in Table 3. These parameters settings are outside of what we might describe as human normal range, since humans tend to learn /z/. For these 796 models we observed again that failure is more likely when decay was less (e.g., .999) and when learning rate is less (e.g., .05), showing how our some of the crossover results are in the same parameter range as other model failures. This indicates that the model does need some learning and some decay to function as humans might be expected to. Again we see the abstraction, engagement threshold and engagement noise are not important to the success. Looking at retrieval threshold completes the picture. We find that a low threshold guaranteed /z/ interpretation (i.e., none of the 796 failures occurred with low threshold). Complementing this result, we found that only 67 of the 796 failures occurred with a high amount of noise.

Table 3

Results for all simulation parameter settings that ended with /z/ recognition.

Model parameter	Low	Medium	High
λ	297	256	243
ρ	282	270	244
δ	264	266	266
Engagement threshold	266	263	267
Engagement threshold noise	267	264	265
Retrieval threshold	0	243	553
Retrieval threshold noise	486	243	67

Together these model results explain /z/ interpretation delayed development as a problem with the learner screening off certain prior experiences, thus incompletely recalling their experiences with /z/ and /s/. When retrieval threshold is low enough, the learner becomes more sensitive to their own prior experiences, reacting to them more appropriately. It appears that a high amount of noise may be sufficient to overcome this problem of a high retrieval threshold. A high amount of noise means that some of the subjects in the sample may have low threshold despite the group, and this might be expected to cause some of the results to favor /z/ when noise is high.

In conclusion, searching the parameter space of the model reveals it behaves sensibly at border conditions, for example, it has no trouble capturing the faster learning of frequent phonemes. In contrast, when an infrequent phoneme presentation occurs the model suggests that it will only be recognized if there is enough experience and if the model parameters are in what we might consider to be the normal human range. Specifically, for the settings we tested the model parameter describing the breadth of memory search for the test trials, the retrieval threshold, had to be sufficiently low so as to be able to benefit from the effect of past experience. At the lowest setting of retrieval threshold learning of /z/ always succeeded. It seemed this screening off of relevant past memories by high retrieval threshold could be occasionally overcome within groups of simulated subjects, if the noise on the retrieval threshold was sufficient. Higher noise leads to some groups of simulated subjects by chance having a low enough threshold so that they also succeed in retrieved discrimination.

These results imply that parameter differences might be a useful tool for exploring and predicting individual differences in subjects. Encouragingly, the model did not show any patterns that would suggest an excess of flexibility in the model, such as if the model had been able to predict /z/ would be learned before /s/. These results suggest that it is plausible to suppose that the parameters map to actual cognitive functions. While we only tested limited ranges of each parameter, the ranges were not excessively narrow and we would similarly expect normal human ability to be also distributed in limited ranges.

Experiment 1A

The results of the previous set of simulations suggest that frequency can have a powerful influence on the emergence of phonemic contrasts. iMinerva is better able to differentiate a phonemic contrast where both contrastive phonemes occur frequently (/d/-/t/) than a phonemic contrast where only one member of the pair occurs frequently (/s/-/z/). These results lead to a straightforward prediction. If iMinerva's simulation of the processes underlying infants' use of phonemic contrasts is accurate, then infants should be able to use more frequent contrasts before they are able to differentiate less frequent contrasts. However, this prediction has not been assessed by previous research. Most prior research has found that infants succeed in responding differentially to phonemic contrasts in a

word–object association task by 17 or 18 months (e.g., Thiessen, 2007; Werker et al., 2002). But all of this research has focused on word-initial stop contrasts in which both phonemes are highly frequent.

In Experiment 1 and Experiment 2, we will examine the developmental progression of the use of the /s/–/z/ contrast. Such an examination is informative in its own right, because infants' use of a fricative contrast has not previously been explored using the Switch task, the most common word–object association task in this area of research (Stager & Werker, 1997; though see Shvachkin, 1973, for a discussion of children's use of fricatives in a different referential task). In addition, assessing infants' use of the /s/–/z/ contrast will allow us to test iMinerva's prediction that infants' use of this contrast should emerge later than the stop consonant contrasts (such as /d/–/t/) typically tested in the Switch task (e.g., Thiessen, 2007; Thiessen & Yee, 2010). But because fricatives have not previously been used in the Switch task, it is important to first demonstrate that the expected inability to take advantage of phonemic contrasts in word–object association tasks is present. This is our goal in Experiment 1. In Experiment 2, we will look to see at what age infants can begin to successfully differentiate between /s/ and /z/ in the Switch task.

Method

Participants

Participants were 25 toddlers between the ages of 17 and 18 months ($M = 17.35$). In order to obtain data from these 25 children, it was necessary to test 30. The additional five participants were excluded for the following reasons: fussing or crying (3), failure to habituate (1), and experimental error (1). According to parental report, all children were free of ear infection at the time of testing, and reported no history of hearing problems.

Stimuli

Participants were exposed to a novel computer-animated object, paired with a novel label. The object was animated, and moved against a black background; this object has been previously used in several variants of the Switch procedure (Stager & Werker, 1997; Thiessen, 2007; Thiessen & Yee, 2010). The object was paired with the label “seer” or “zeer” (IPA: /sir/ and /zir/; the identity of the label was counterbalanced across infants). The labels were recorded by a female native speaker of English in an infant-directed prosody. Each label was repeated for as long as the object remained on screen, with pauses of 1.4 s between repetitions.

Procedure

This experiment used a habituation procedure identical to that used in Thiessen (2007). Participants, seated on a parent's lap in a sound-attenuated room, controlled the duration of the presentation of the stimuli by the length of their gaze at a central monitor. To eliminate bias, parents wore headphones, and the experimenters sat in an adjacent room, blind to the nature of the stimulus being presented. The experimenter coded the duration of the child's looking time online, using a Macintosh running

the Habit program for OS X (Cohen, Atkinson, & Chaput, 2004). After the child reached the habituation criterion (looking time less than 50% of the average of the first three trials), six test trials were presented.

The child controlled the duration of stimulus presentation, during both the habituation and test phases, by gazing at a 32" video monitor 150 cm in front of their seated position. The child's attention was attracted to the monitor by a colorful video of Winnie the Pooh coupled with a recorded verbal encouragement. Once the experimenter believed the child's attention was fixated on the monitor, stimulus presentation was initiated. An object then appeared on the screen and the speakers adjacent to the monitor began to repeat label associated with that object. The stimulus presentation continued until the child looked away for more than 1 s, or until the child had gazed at the monitor for 20 s (the maximum time allowed per trial). The video of Winnie the Pooh appeared at the end of each trial to recapture the child's attention.

The experiment began with two familiarization trials, in which participants saw a nonsense object paired with the word “neem.” The familiarization trials were used to help participants become accustomed to the pairing of audio and visual stimuli. Once these trials were finished, the habituation phase began.

During the habituation trials, the novel object appeared on screen, and was paired with the label (either “seer” or “zeer,” counterbalanced across infants); for all children, the label was an identical token repeated multiple times. The object moved onscreen while the label was repeated, with pauses of 1.4 s between each repetition. The object and the label were presented until the child looked away from the monitor for 1.5 s, at which point the attention-getting stimulus reappeared. Looking times to each trial were calculated in real time, and the habituation trials continued in random order until the child met the habituation criterion: average looking time for three consecutive trials that fell below 50% of their looking time to the first three habituation trials.

Once the child met the habituation criterion the 6 test trials began. In all test trials, the child saw the object from the habituation phase on the monitor. There were two kinds of test trials: Same and Switch trials. In the Same trials, the children heard the syllable that they had previously heard paired with the object. In the Switch trials, the object was paired with a minimal pair of the label infants had heard in the habituation phase. For infants exposed to the object paired with “seer,” the Switch trial paired the object with the novel label “zeer.” The opposite was true for infants exposed to “zeer” during the habituation phase.

Same and Switch trials alternated, and the nature of the initial test trial was counterbalanced across participants. As in the habituation trials, the object stayed on the screen, and the label continued to repeat, for as long as the participant continued to look at the monitor.

Results and discussion

On average, children habituated in 13.4 trials. This is a somewhat longer habituation phase than has been seen in prior experiments with stop consonants (see Table 4),

Table 4

Comparison of the results from Experiment 1 with prior Switch Task results (Werker et al., 2002). Times in parentheses indicate standard error of the mean. Bracketed descriptions “ignore” and “detect” describe whether infants responded to the phonemic contrast.

Experiment	Phonemic Contrast	Participant Age	Looking time to Same	Looking time to Switch
Thiessen and Yee (2010), Experiment 1	/d/-t/ [ignore]	14.5–15.5 months	5.9 (0.7)	5.7 (0.7)
Werker et al. (2002), Experiment 2	/b/-d/ [ignore]	14–15 months	8.6	8.5
Werker et al. (2002), Experiment 3	/b/-d/ [detect]	17–18 months	8.4	11.8
Experiment 1A	/s/-z/ [ignore]	17–18 months	5.6 (0.6)	5.5 (0.5)
Experiment 2	/s/-z/ [ignore]	19–20.5 months	5.2 (0.7)	5.1 (0.6)
Werker et al. (2002), Experiment 1	/b/-d/ [detect]	20–21 months	9.5	12.5
Experiment 2	/s/-z/ [detect]	24–25.5 months	4.2 (0.5)	6.3 (0.8)

perhaps due to children’s differential familiarity with stop and fricative consonants. There was no significant difference in the number of trials children required to habituate to *seer* ($M = 13$) or *zeer* ($M = 13.9$), $t(23) < 1$, *n.s.* More importantly, children’s preference for novel vs. familiar test trials did not differ as a function of which item they heard during the habituation phase: $t(23) < 1$, *n.s.* Therefore, for all subsequent analyses, participants were grouped together, regardless of which label they heard during the habituation phase.

Our primary question was whether infants in the test phase responded differentially to trials in which the novel object was paired with its original label (Same trials), vs. trials where the object was paired with a minimal pair of that label (Switch trials). On average, children’s looking time to Same trials was 5.6 s ($SE = 0.6$). Their mean looking time to Switch trials was 5.5 s ($SE = 0.5$). 24 out of 25 children showed a difference between Same and Switch trials of less than 1 s. A two-tailed *t*-test (all *t*-tests reported are two-tailed) indicated that this difference was not significant: $t(24) < 1$, *n.s.*

These results are consistent with prior experiments using the Switch paradigm, which have demonstrated that children often fail to respond differentially to lexical forms differentiated by only a single phoneme (e.g., Pater et al., 2004). Indeed, although we investigated a novel fricative contrast, the data from these children are strikingly similar to prior results from our laboratory with stop consonants (see Table A1). This experiment demonstrates that children’s difficulty with minimal pair contrast extends to fricative contrasts.

Experiment 1B

The results of Experiment 1A indicate that 18-month-old infants have difficulty using the /s/-z/ distinction in a word-object association task. This is consistent with the general claim of the acquired distinctiveness account, supported by the iMinerva simulations, that less frequent contrasts should be acquired later. However, there is an alternative explanation for these results, which is that the /s/-z/ contrast is indistinguishable to children for perceptual reasons rather than experiential ones. For example,

it may be the case that children have difficulty hearing the voicing contrast in the context of a fricative consonant, while they can perceive it more easily in the context of a stop consonant like /d/ or /t/.

To address this issue, in Experiment 1B we presented children with a simpler version of the Switch task, one in which the labels are presented absent any objects. Prior work has demonstrated that in the absence of a to-be-labeled object, children of this age are able to discriminate phonemic contrasts that they fail to use in a word-object association task (e.g., Stager & Werker, 1997; Thiessen, 2007). If children can perceive the difference between /s/ and /z/, they should succeed in discriminating them in Experiment 1B, where there is no labeling component to the task. By contrast, if children’s failure in Experiment 1A stems from some perceptual difficulty in discriminating the /s/-z/ contrast, they should also fail to discriminate between the phonemes in Experiment 1B.

Participants

Participants were 25 toddlers between the ages of 17 and 18 months ($M = 17.2$). In order to obtain data from these 25 children, it was necessary to test 27. The additional two participants were excluded for the following reasons: failure to habituate (1), and experimental error (1). According to parental report, all children were free of ear infection at the time of testing, and reported no history of hearing problems.

Stimuli

The acoustic stimuli were identical to those used in Experiment 1A. Rather than being paired with an animated object, however, the syllables were paired with a static image of a monochromatic checkerboard during both the habituation and the test phase (cf. Stager & Werker, 1997; Thiessen, 2007).

Procedure

The procedure was identical to that used in Experiment 1A.

Results and discussion

On average, children habituated in 8.9 trials. There was no significant difference in the number of trials children required to habituate to *seer* ($M = 8.8$) or *zeer* ($M = 9$), $t(23) < 1$, *n.s.* After habituation, children's preference for novel vs. familiar test trials did not differ as a function of which item they heard during the habituation phase: $t(23) < 1$, *n.s.* Therefore, for all subsequent analyses, participants were grouped together, regardless of which label they heard during the habituation phase.

On average, children's looking time to Same trials was 3.5 s ($SE = 0.5$). Their mean looking time to Switch trials was 5.6 s ($SE = 0.7$). 22 out of 25 infants looked longer to Switch trials than to Same trials. A two-tailed *t*-test indicated that this difference was significant: $t(24) = 3.1$, $p < .05$. This result demonstrates that children are able to perceive the /s-/z/ contrast. The difficulty in responding differentially to this contrast is selectively present in a word-object association task.

Experiment 2

The results of Experiment 1 demonstrate that the /s-/z/ phonemic contrast results in the same kind of difficulty in the Switch task as do more common (and commonly tested) stop consonants: once infants have associated an object with the label /sir/, they accept /zir/ as an interchangeable label for the same object. Prior research has shown that for stop consonants, children can succeed in the Switch task by 18 months (e.g., Thiessen, 2007; Thiessen & Yee, 2010; Werker et al., 2002). One possible explanation for this is that young children have limited processing capacity, which prevents them from encoding the label in the context of a label-object matching task. From this perspective, the reason why children succeed once they have reached a threshold age is that they have exceeded some critical capacity threshold (for discussion, see Werker et al., 2002). If this account is correct, we would expect that 20-month-olds would succeed with a wide range of consonants, not just stop consonants, because they have the necessary cognitive capacity to represent both the auditory and visual information necessary to succeed in the Switch task.

Alternatively, we have proposed that the developmental change in children's use of phonemic contrasts is not related to overall gains in capacity, but to children's increasing familiarity with phonemes due to an expanding vocabulary (Thiessen, 2007). If this is the case, children should succeed at 18–20 months with phonemic contrasts with which they are more familiar, but potentially fail with phonemic contrasts with which they are less familiar. This is consistent with the prediction emerging from iMinerva that children's differentiation of phonemic contrasts depends on the number of different lexical contexts in which the contrast occurs. Because /s-/z/ occurs much less frequently in word-initial position than the stop consonants previously tested in the Switch paradigm, we hypothesize that 20 month-olds will fail to respond differentially to this contrast. To test this hypothesis, we used

the stimuli and procedure from Experiment 1 to test both 20- and 25-month-old children on their ability to use the /s-/z/ contrast in a word-object association task.

Method

Participants

Participants were two groups of 25 toddlers, one group between the ages of 19.0 and 20.5 months ($M = 19.8$), and the other group between the ages of 24.0 and 25.5 months ($M = 24.9$). In order to obtain data from 25 20-month-olds, it was necessary to test 29 participants. The additional four participants were excluded for fussing or crying (3) or failure to habituate (1). In order to obtain data from 25 25-month-olds, it was necessary to test 33 participants. The additional eight participants were excluded for fussing or crying (6) and failure to habituate (2). According to parental report, all children were free of ear infection at the time of testing, and reported no history of hearing problems. All participants were given a questionnaire for parents to complete prior to visiting the lab. This questionnaire contained all of the /s/- and /z/-initial words found on the MacArthur CDI: Words and Sentences form. As in the original MacArthur, for each word parents were asked to indicate whether their child comprehended and/or produced the word.

Stimuli and procedure

The stimuli and procedure were identical to those used in Experiment 1A.

Results and discussion

20-month-olds

On average, 20-month-old children habituated in 7.6 trials. There was no significant difference in the number of trials children required to habituate to *seer* ($M = 7.8$) or *zeer* ($M = 7.5$), $t(23) < 1$, *n.s.* Subsequent to habituation, children's preference for novel vs. familiar test trials did not differ as a function of which item they heard during the habituation phase: $t(23) < 1$, *n.s.* Therefore, for all subsequent analyses, participants were grouped together, regardless of which label they heard during the habituation phase.

Like the younger children in Experiment 1A, 20-month-old children failed to respond differentially to Same and Switch trials. 21 out of 25 children showed looking times to Same and Switch trials that were within 1 s of each other. On average, children's looking time to Same trials was 5.2 s ($SE = 0.7$). Their mean looking time to Switch trials was 5.1 s ($SE = 0.6$). A two-tailed *t*-test indicated that this difference was not significant: $t(24) < 1$, *n.s.* These results are consistent with iMinerva's predictions: while 20-month-olds can succeed with more frequent phonemic contrasts (e.g., Thiessen & Yee, 2010; Werker et al., 2002), they do not respond differentially to the less frequent /s-/z/ contrast.

25-month-olds

On average, 25-month-old children habituated in 8.5 trials. There was no significant difference in the number

of trials children required to habituate to *seer* ($M = 8.0$) or *zeer* ($M = 8.9$), $t(23) < 1$, *n.s.* Subsequent to habituation, children's preference for novel vs. familiar test trials did not differ as a function of which item they heard during the habituation phase: $t(23) < 1$, *n.s.* Therefore, for all subsequent analyses, participants were grouped together, regardless of which label they heard during the habituation phase.

Unlike 20-month-olds, 25-month-olds responded differentially to the */s/-/z/* distinction. 16 out of 25 children looked longer to Switch trials than to Same trials. On average, children's looking time to Same trials was 4.2 s ($SE = 0.5$), while their looking time to Switch trials was 6.3 s ($SE = 0.8$). This difference was significant, $t(24) = 2.4$, $p < .05$. These results are consistent with iMinerva's prediction that use of the */s/-/z/* contrast will take longer to emerge because children experience it less frequently.

Note that at least in one respect, however, the iMinerva simulations are potentially discrepant from the behavioral results. iMinerva is even more sensitive to phoneme frequency than our behavioral results, in that iMinerva shows an asymmetry in its response to the less frequent */z/* and the more frequent */s/*; at some points in learning, iMinerva discriminates an */s/-probe* from */z/*, but does not distinguish a */z/-probe* from */s/*. By contrast, our 17–20 month olds (who failed to respond differentially to either contrast) and our 25-month-olds (who responded differentially to both) showed no such asymmetry. One possibility is that there is an intermediate age group (e.g., 23-month-olds) who would succeed in the task when trained on */s/* but fail on when trained on */z/*; the 25-month-olds may simply be too old to show iMinerva's predicted asymmetry. Another (not mutually exclusive) possibility is that our experiments were insensitive to the predicted asymmetry. One reason for this is that while we counterbalanced */s/* and */z/* initial words during familiarization, this resulted in a smaller sample size for either kind of word, so the experiments may have been insensitive to any differences between the phonemes. The Switch task itself may also be a poor instrument for detecting relatively subtle differences between phonemes. Recall that in the Switch task, the child simply has to respond differentially to the novel trial. It may be the case that even a weaker representation of */z/* is sufficient for the child to detect that the novel stimulus ("seer") is not identical to the familiarized stimulus ("zeer"). If this is the case, experiments using a more sensitive measure, such as reaction time (e.g., Yoshida et al., 2009), may detect differences in children's response to */s/-initial* and */z/-initial* words.

A related prediction of the acquired distinctiveness account, and the iMinerva simulations, is that individual differences in children's familiarity with the */s/-/z/* contrast should be related to individual differences in their response to the contrast. That is, children who have more experience with */s/* and */z/* in distinct lexical contexts should be more likely to make use of the contrast in a word-object association task. To explore this prediction, we assessed how children's parent-reported familiarity with */s/-/z/* words predicted the degree to which children dishabituated. If the simulations reported above are correct, there should be a positive correlation between

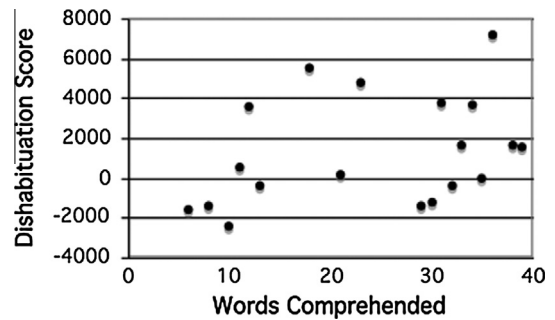


Fig. 2. Scatterplot of individual 25-month-old children's degree of dishabituation in Experiment 2 related to the number of */s/* and */z/* words children comprehend, via parental report on a subset of the MacArthur-Bates CDI. Negative values on the Y-axis (dishabituation score) indicate a preference for the Same test item; positive values indicate a preference for the Switch test item.

reported vocabulary size and children's use of the */s/-/z/* contrast. If the protracted emergence of the */s/-/z/* contrast is due to some acoustic idiosyncrasy, there is no reason to expect such a correlation.

Twenty-one of 25 participants' parents completed the vocabulary questionnaire, which assessed their knowledge of */s/-* and */z/-initial* words (note that only */s/* and */z/* words were assessed, as the full MacArthur CDI took too long for parents to complete given the number of vocabulary items children of this age were familiar with). Comprehension scores ranged from 0 to 71 ($M = 30.1$, $SD = 17.1$). We correlated the number of words a child comprehended with a difference score (looking time to Switch trial – looking time to Same trial) representing their degree of dishabituation. There was a significant positive correlation between these scores, $r = .44$, $p < .05$ (see Fig. 2). This correlation indicates that the more */s/-/z/* initial words a child knew, the greater the magnitude of their dishabituation to Switch trials. This is consistent with iMinerva's prediction that more experience with */s/* and */z/* in lexical contexts should make children better able to differentiate between them.

General discussion

Despite the rapid gains infants make in adapting to the phonemic structure of their native language in the first year of life (e.g., Werker & Tees, 1984), children still struggle to use these phonemic contrasts in the Switch task well into the second year of life (e.g., Pater et al., 2004; Stager & Werker, 1997; Thiessen, 2007; Thiessen & Yee, 2010). We propose that the later emergence of the ability to use phonemic contrasts in a word-object association task is that not all perceptual contrasts are lexical contrasts. For example, indexical information also varies across lexical exemplars, and is readily perceptible to infants, but these indexical differences do not indicate a difference in lexical category.

The acquired distinctiveness account suggests an avenue via which infants can identify which perceptual distinctions are relevant in differentiating lexical items: by learning from the distribution in the input (for a related

argument, see Apfelbaum & McMurray, 2011). When two phonemes occur in distinct lexical contexts (such as /d/ and /t/ in “doggy” and “teddy”), they become more distinctiveness and more readily available for use. Because infants know relatively few minimal pair words (Caselli et al., 1995), most of the lexical items that infants learn should serve to make phonemic contrasts more distinctive. Therefore, as infants become familiar with more lexical forms, they should become better able to use the phonemes embedded in those forms.

This account differs from prior computational accounts of the acquisition of phonetic contrasts in a variety of ways. First, unlike most computational accounts of the acquisition of phonetic categories (e.g., Vallabha, McClelland, Pons, Werker, & Amano, 2007), it is not focused on how infants learn to discriminate among phonetic exemplars to identify category structure; it is focused on how infants learn to use these distinctions. Merely being able to make a perceptual distinction is not sufficient for knowing that a distinction is informative about a difference in meaning, as many perceptual distinctions (e.g., speaker identity) are uninformative with respect to lexical meaning. The distributional learning account is unique in that it focuses on the context in which these phonemes occur, and the frequency with which children receive that contextual information, to predict when and how children treat a phonetic contrast as semantically contrastive. Second, iMinerva invokes a novel (at least, with respect to phonetic learning) mechanism to explain the role of variability in the development of phonetic contrasts. The most compelling prior computational account of the role of variability in phonetic learning is an associative learning account (Apfelbaum & McMurray, 2011). On this account, when children experience a phonetic contrast in different contexts (e.g., produced by different speakers), they downweight attention to the variable dimensions, to focus on the information (the phonetic contrast) that is predictive. By contrast, in our computational approach, the variability becomes an essential part of the representation of the contrast. When a child hears /d/ in one set of contexts, and /t/ in a different set of contexts, these different contexts are automatically “called to mind” in subsequent experiences with those phonemes, and serve to pull them apart in representational space. While we have focused on lexical context, an intriguing possibility of this approach is that other contexts – such as the object with which an object is paired (e.g., Yeung & Werker, 2009), or the speakers who produce the contrasts – may be part of that representational space.

The current results provide two novel pieces of evidence in support of the acquired distinctiveness account. First, our simulations suggest that natural input (at least in English) provides the kind of distributional structure necessary for children to learn to use phonemic categories via acquired distinctiveness. Across those lexical items with which English-learning children are familiar, phonemes tend to occur in distinct lexical contexts, and these contexts provide enough evidence for a simulated learner to become better able to differentiate phonemic contrasts as a function of increasing familiarity with a set of lexical items. This suggests that the developing lexicon may play a crucial role in driving children’s increasing ability to

make use of phonemic contrasts in a word–object association task (cf. Werker et al., 2002). As children learn more words, they acquire more evidence differentiating the representations of similar phonemes, because those phonemes tend to occur in different lexical frames. This suggests that children’s failures in the Switch task are not due to capacity limitations or a lack of social support, but are instead informative about their developing phonemic representations.

Necessarily, simulations like iMinerva provide only an existence proof that learning from the input is possible, not a demonstration that humans learn in the same way as the model. However, the acquired distinctiveness framework, instantiated in iMinerva, makes a novel prediction about human learning: the more frequent phonemes should emerge sooner than less frequent phonemes. Note that this is not the only possible developmental account. For example, Werker et al. (2002) suggested a capacity account in which failures in the Switch task could be explained via children’s lack of sufficient capacity to encode and maintain acoustic detail in the context of a word–object association task. This account predicts that once children have reached a certain threshold of capacity, they should be able to succeed in the Switch task, regardless of which contrast they are asked to use. Indeed, this is a common implicit assumption in the literature. On the basis of children’s success at 18–20 months with stop consonants (like /b/–/d/), it is often assumed that children can succeed with all consonantal distinctions in the Switch task by this age.

As our results demonstrate, this is not the case. Even 20-month-old infants fail to make use the /s/–/z/ distinction in a word–object association task. To our knowledge, this is the first demonstration of failure in the Switch task at this older age. Experiment 1B indicates that this is not due to an inability to discriminate the /s/–/z/ contrast. Rather, we suggest that it is due to the fact that the /s/–/z/ contrast is much less frequent than the stop consonants that have been previously used in the Switch task (e.g., Pater et al., 2004; Stager & Werker, 1997; Thiessen, 2007; Thiessen & Yee, 2010). Because the contrast is less frequent, it takes children longer to begin to use it productively. This should not be taken to mean that frequency is the only reason that children use fricative contrasts later than stop consonants. A variety of empirical work suggests that fricatives may present a more difficult perceptual challenge than common stop consonants (e.g., Abbs & Minifie, 1969). Indeed, the lower frequency of fricative contrasts – especially in infant directed speech – may be the result of linguistic adaptation to the kinds of contrasts that infants find easier to perceive. Instead, these results suggest that frequency and context play a role, but certainly not the only role, in making a contrast accessible to the semantic system. An important avenue for future work will be to replicate this role of frequency and context across a wider array of contrasts to begin to tease apart the independent contributions of distributional information and perceptual information.

What these results suggest, however, is that failure in the Switch task cannot solely be ascribed to infants’ and young children’s inability to represent perceptual detail

(for related arguments, see Swingley & Aslin, 2000, 2002; Yoshida et al., 2009). Instead, performance in the Switch task reflects how much evidence infants have acquired that a phonetic contrast reflects differences at the lexical level. The more information infants have acquired indicating that two phonemes are distributed in two different sets of lexical items, the more like they are to use the contrast between those phonemes. Because the /s/-/z/ contrast is relatively rare, it takes longer to emerge productively. Furthermore, consistent with the acquired distinctiveness account, when children are on the cusp of using a contrast in the Switch task it is possible to predict how well they will do as a function of their experience with that particular contrast. In Experiment 2, the children who had more distributional information about /s/ and /z/ (that is, who knew more /s/ and /z/-initial words) were more likely to succeed in the task.

Taken together, these results provide strong support for the acquired distinctiveness account. They demonstrate that natural language provides the necessary distributions, and they demonstrate that the degree of experience that infants have with those distributions is predictive of their performance in the Switch task. Furthermore, these results suggest that iMinerva is a potentially useful tool for simulating distributional learning tasks, insofar as it makes novel, testable predictions about development in response to these patterns.

Acknowledgments

This research was funded by an NSF career award to Erik D. Thiessen, NSF BCS-0642415. We would like to thank Meagan Yee, Teresa Pegors, and Ashley Episcopo for help in the conduct of this research, as well as the parents of all of the infants who participated.

Appendix A

Our model is called interpretative Minerva (iMinerva) because it extends Hintzman's MINERVA 2 model (Goldinger, 1998; Hintzman, 1988) with the capability to maintain not only examples in memory, but also interpretations of examples. In the model, each new example that the learner encounters is compared with prior examples to determine the similarity with these prior examples. This comparison is proposed to be an automatic process of human cognition according to the model and corresponds to learner's basic ability to interpret new experience through the lens of old experience (learning is constructive). If multiple prior examples are similar, the learner selects the strongest of them. Assuming a prior example is selected; the learner engages with the prior example to modify it and create an interpretation. This synthetic interpretation is then recorded as a new memory item.

Each interpretation the learner forms functions like an example according to the model, so that once interpretations are formed, they are themselves engaged with by new examples. In this way, interpretations are like concepts that originally develop from a perceptual experience, but then get increasingly divorced from perceptual experi-

ences as multiple perceptual encounters shape conceptual learning (Sloutsky, 2009). Learning in the model occurs as simple adjustment of the prior example or interpretation trace (or more simply, the memory) using an additive learning rule to determine how prior memories grow by a proportion of the new similar example. This mechanism creates a more general representation by blending together items that originally share some similarity. If the new example has a feature that is different than an old example, despite being similar overall, this learning averages the old feature with a portion of the new feature to reduce the strength of the feature that is different in the interpretation.

Interpretation in this way is a mechanism for prototype creation in the model. If, for example, prior example *A* and new example *B* are found to be similar, they may be engaged. In this case, if *A* has feature 1 = -1 and *B* has feature 1 = 1, then the interpretation created will show a feature 1 that is moved from -1 closer to 1 during learning. While this means that the interpretation is more general, that generality will still not generalize well because feature 1 is still included in the interpretation. For this reason, if example *A* is repeated it will still match with prior example *A* better than the interpretation, while if example *B* is repeated it will also match with prior example *B* better than the interpretation.

To resolve this issue we introduce a very simple abstraction mechanism that removes features from interpretations when those features are some fraction of the maximum absolute feature strength. This abstraction mechanism seems a natural addition to the system, since salience is something that the brain seems to encode directly (Gottlieb, 2007) and our abstraction mechanism is inherently a mechanism that abstracts away less salient features. Not only do we find evidence for salience related information at the level of parietal cortex activity, but we also see that there appear to be cognitive benefits of using prototypes (Winkelman, Halberstadt, Fazendeiro, & Catty, 2006). Furthermore, there are very good reasons to believe that humans are limited in their focus of attention (Miller, 1956), and so, interpretations must necessarily become abstract because of the lack of an ability to attend to all the features in a stimulus and learn them all each time the stimulus is encountered in the environment.

A.1. Prior specification

In iMinerva memory traces are represented as vectors of real numbers where some features may be null values. In contrast, MINERVA 2 requires 1, 0 or -1 values. This change in the feature coding provides a representation that allows us to capture both the strength and durability of an interpretation as the absolute value of a feature's strength. In this formalism, 0 comes to mean either a weak or very equivocal feature, and in either case, we allow features to transition to a null value when they are near 0. This mechanism (abstraction, described below) allows us to represent salience as a binary quantity that depends upon feature strength. This binary salience was a simplification of more complex alternatives that would have required feature salience as a continuous quantity for each feature.

Table A1 shows how each syllable's consonants and vowels were represented.

Since we use this representation format for our interpretations, we also needed a new similarity function, as MINERA 2 simply uses the weighted average of feature agreement. Because our features now represent the strength of each feature in the interpretation, we are no longer looking for the mere binary agreement of features, but rather how the pattern of strengths in the exemplar is similar to the pattern of strengths in the interpretation. Because of this we have adopted a well-established measure, cosine similarity, because cosine similarity compares the magnitude pattern while MINERVA 2 similarity only weights yes or no agreement of features. In addition to handling the magnitudes, cosines similarity has a long history of use in text classification (Salton, 1989). Eq. (1) below shows the cosine similarity function. Like Hintzman's MINERVA 2, we compress the results of this metric to compute similarity by cubing the raw cosine similarities to increase dispersion among the values obtained. Hintzman refers to the cubed value as representing activation the activation of a memory by a similar probe (Goldinger, 1998; Hintzman, 1988). The calculation of our similarity measure, cosine similarity cubed, is shown in Eq. (1). Furthermore, we have modified this traditional equation such that if a feature is missing (has been abstracted away in the case of interpretations, or was not present in the stimuli in

the case of examples) from either trace *A* or trace *B*, that feature is ignored in the computation.

$$\text{Similarity}_{A,B} = \left(\frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \right)^3 \quad (1)$$

iMinerva processes the example input stream in discrete time steps representing the cognitive cycle (Madl, Baars, & Franklin, 2011) for each word encounter. Each new example is compared with all prior memory traces (both interpretations and examples) to see if the engagement threshold parameter is exceeded. If the engagement threshold is exceeded, it indicates that the learner notices the match(s) with prior stimuli. If nothing is matched, no interpretation is formed. While the underlying comparison process is assumed to unfold over time, the model operationalises the outcome by a simple "max similarity rule." Eq. (3) shows how a prior trace, *A*, accumulates a portion of the strength of *B* when the similarity of *B* exceeds engagement threshold and is the maximum similarity trace that exceeds engagement threshold. The learning rate for this accumulation is represented by λ .

Given probe example *A*, and X_n ,
the *n*th item from the set of *i* memories
if $\text{Similarity}_{A,X_n} > \text{threshold}$
and
 $\text{Similarity}_{A,X_n} = \text{argmax}(\text{Similarity}_{A,X_{1..i}})$
then
create a new memory, $X_{i+1} = X_n + \lambda A$ (2)

This learning process is both strengthening (Eq. (2)) and abstractive. The abstractive component is captured in Eq. (4) with specifies that given any feature in the interpretation, it will be removed if it is weak relative to the maximal feature. This means that the absolute value of the strength of any feature must exceed the engagement threshold for that feature to be retained. Eq. (3) describes how each feature must exceed this criterion. Eq. (3) uses the ρ parameter which is the fraction of the maximum absolute value that must be exceeded to retain a feature, otherwise that feature is set to null.

Given X_m , where *m* indexes features 1...*j* for memory *X*
if $|X_m| < \rho(\text{argmax}|X_{1..j}|)$
set $X_m = \text{null}$ (3)

Finally, we included forgetting in the model, which we simulate with simple exponential decay. While exponential decay may be a less accurate than power law or other functions in modeling forgetting (Rubin & Wenzel, 1996), in this model where forgetting is not a key factor, this decay mechanism adds plausibility to the model because it illustrates how memory traces are lost and why the model does not need to be concerned about the criticism that storing unlimited examples is implausible. The model is explicitly limited in the examples it can store because old examples eventually decay to the point they are never engaged, and are therefore essentially deleted. Further, we

Table A1

Descriptions of the feature meanings for the phoneme vectors used by iMinerva.

Type	Index	Description
C	1	voicing
C	2	manner: stop
C	3	manner: fricative
C	4	manner: affricate
C	5	manner: nasal
C	6	manner: lateral
C	7	manner: retroflex
C	8	manner: glide
C	9	place: labial
C	10	place: interdental
C	11	place: alveolar
C	12	place: alveopalatal
C	13	place: velar
C	14	place: glottal
V	1	vowel [bead]
V	2	vowel [bid]
V	3	vowel [bayed]
V	4	vowel [bed]
V	5	vowel [bad]
V	6	vowel [bod]
V	7	vowel [foot]
V	8	vowel [bode]
V	9	vowel [bood]
V	10	vowel [bud]
V	11	vowel [bird]
V	12	vowel [bide]
V	13	vowel [boy]
V	14	vowel [bore]
V	15	vowel [blouse]
V	16	vowel [bear]
V	17	vowel [beer]
V	18	vowel [bard]

should not that we do not have a “theory of decay” other than to argue that interference appears to be a primary cause of decay (Pavlik & Anderson, 2005). Eq. (4) shows decay in the model for some example feature vector, N .

Given $X_{m,t}$, where m indexes features $1 \dots j$
for memory item X at trial t

$$X_{m,t+1} = \delta X_{m,t} \quad (4)$$

A.2. New mechanisms

Similarity figures into our analysis both in running the model as it makes interpretations, but also in the more basic process of computing the memory echo. While we did not discuss it in our previous paper, the overall similarity our model retains with Hintzman’s MINERVA 2 means that we can use the MINERVA 2 memory echo intensity and content functions as a way to measure the memory structures created under different conditions. Echo intensity is simply defined as the sum of the similarity (Eq. (1)) of the probe to all of the prior memories. Echo content is created by multiplying each memory’s similarity (a measure that determines how “activated” a memory becomes by a probe) by the memory vector, and then in a second step sums all these individual activated traces to get the summed echo content for a specific probe.

Using this standard method (Goldinger, 1998) to compute echo content we considered how to compare the echo contents to different probes to see what is recalled under various circumstance by the model. We desired to test the idea that the distributional pattern differences in the echo content allow the infant to discriminate the probe sounds. We propose that the “echo content” allows infants to attach different labels to sounds. To the extent that two sounds cue memory patterns that are very similar, they will not be differentiated. Conversely, if the overall pattern of features recalled for two different probes differs, they will be considered as members of unique classes. However, we also realized that strength of the memory would need to play a role, and we needed a recall equation that would represent both strength and match of prior memories. We did not get that from MINERVA 2.

In the paper we describe in detail how we used the echo content function to create a new discriminative recall strength equation that asks the question: given the echo contents of the possible probes, which echo is most likely to be the identified given a particular probe. The means that the echo contents to the possible probes are used as the model’s measurement of the categories in memory.

In the paper this meant that for each simulation we computed 4 echo content vectors that were compared to see which was identified (i.e., most active) for each probe. We do this with Eq. (5), where we first normalize the activation similarity of the probe with the possible memories (since the correct memory echo contents is a derived from the probe, this normalizes that value to 1, and other values are less than 1, generally positive).

Following this step we take this value to the exponent of the average absolute value of the echo contents, which means that strong memories will cause increased discrimination by increasing dispersion of values less than 1. Finally the value was also multiplied by the average absolute value of the echo contents, which means that strong memories will tend to dominate recall. As detailed in the body of the paper, as practice accumulates, at first the multiplicative term dominates, meaning frequency of prior learning controls recall reducing discrimination, but later on, the exponential “resonance” term dominates as stronger memories resonate and allow more opportunities for incorrect match rejection, thus improving discrimination.

Given probe example A , and echo content from probe x , Q_x :

$$\frac{\sum_{i=1}^n |Q_x|}{n} \left(\frac{\text{Cosine}_{A,Q_x}}{\text{Cosine}_{A,Q_a}} \right)^{\frac{\sum_{i=1}^n |Q_x|}{n}} \quad (5)$$

Additionally, for the new simulation in this paper we changed the echo content function that feeds into Eq. (2), by only including the echoes from memories that were above some retrieval threshold similarity with the probe. This recall breadth parameter was found to control the speed of discrimination from Eq. (5), such that more broad recall results in better discrimination, due to increased resonance from the strong memory signal. We find it plausible to suppose that in addition to improvement in discrimination due to learning, shown in the paper, the improved discrimination in children may come from reductions in the retrieval threshold (increases in the breadth of recall) as a child’s brain develops, and we could use our retrieval threshold parameter as one way to model this development. In other words, our model can be easily used to suggest the combination of maturation and frequency of phonemes experienced contribute to discrimination growth across infancy. However, it is important to point out that other mechanisms like a higher learning rate could also be used to capture this improved discrimination capability we presume is gained as children mature.

Table A2 shows the parameters across the models, which are discussed in the paper body where appropriate.

Table A2

Parameter values used in the simulations.

Parameter	Represents	Simulation	Search
λ	Learning rate	.2	.05, .15, .25
ρ	Abstraction proportion	.1	.05, .15, .25
δ	Decay rate	.998	.997, .998, .999
Engagement threshold	Engagement threshold	.75	.65, .80, .90
Engagement threshold noise	SD of engagement threshold	.05	.05, .15, .25
Retrieval threshold	Retrieval threshold	.425	.325, .475, .625
Retrieval threshold noise	SD of retrieval threshold	.05	.05, .15, .25

The model above is highly simplified to clarify explanation, but we argue that it captures the basic process of general exemplar learning, prototype extraction and prototype recall as it occurs in learners without complex language. We do not argue that the model above is correct or complete, merely that it adds to our understanding by showing a minimal set of principles that can achieve the learning of phonemic discrimination from words. It seems likely that infants have minimal ways to direct the above processes, and that the cycle of experience, learning, and abstraction is driven by physical needs or by attraction to similarities in the environment (particularly similarities to items that were associated with reward in the past). No doubt, humans become quite skilled at guiding this cycle and sculpting their learning as their capabilities for action grow and they develop complex symbolic representations and goal structures.

References

- Abbs, M. S., & Minifie, F. D. (1969). Effect of acoustic cues in fricatives on perceptual confusions in preschool children. *Journal of the Acoustical Society of America*, 46, 1535–1542.
- Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, 46, 155–182.
- Apfelbaum, K., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, 35, 1105–1138.
- Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, 10, 159–199.
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighborhoods in young children's lexicons. *Journal of Child Language*, 17, 205–215.
- Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighborhoods in young children's receptive vocabularies. *Journal of Child Language*, 22, 727–735.
- Coady, J. A., & Aslin, R. N. (2003). Phonological neighborhoods in the developing lexicon. *Journal of Child Language*, 30, 441–469.
- Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). *Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0)*. Austin: University of Texas.
- Fennell, C., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, 81, 1376–1383.
- Fenson, F., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., ... Reilly, J. S. (2002). *MacArthur communicative development inventories: User's guide and technical manual*. Baltimore: Paul H. Brookes.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- Gottlieb, J. (2007). From thought to action: The parietal cortex as a bridge between perception, action, and cognition. *Neuron*, 53(1), 9–16 [10.1016/j.neuron.2006.12.009].
- Hall, G. (1991). *Perceptual and associative learning*. Oxford: Clarendon Press.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95, 528–551.
- Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 1143–1154.
- MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, 31, 883–914.
- Madl, T., Baars, B. J., & Franklin, S. (2011). The timing of the cognitive cycle. *PLoS ONE*, 6(4), e14803. <http://dx.doi.org/10.1371/journal.pone.0014803>.
- McClelland, J. L., & Plaut, D. C. (1999). Does generalization in infant learning implicate abstract algebraic-like rules? *Trends in Cognitive Science*, 3, 166–168.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Nissen, S. L., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *Journal of the Acoustical Society of America*, 118, 2570–2578.
- Pater, J., Stager, C. L., & Werker, J. F. (2004). The lexical acquisition of phonological contrasts. *Language*, 80, 361–379.
- Pavlik, P. I., Jr., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, 29, 559–586.
- Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, 103, 734–760.
- Salton, G. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Addison-Wesley Longman Publishing Co., Inc.
- Shvachkin, N. Kh. (1973). The development of phonemic speech perception in early childhood. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development* (pp. 91–127). New York: Holt, Rinehart, & Winston (Original work published in 1948).
- Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, 106, 833–870.
- Sloutsky, V. M. (2009). *From perceptual categories to concepts: What develops?*
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than word-learning tasks. *Nature*, 388, 381–382.
- Swingle, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B*, 364, 3617–3622.
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76, 147–166.
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13, 480–484.
- Swingle, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, 54, 99–132.
- Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, 56, 16–34.
- Thiessen, E. D. (2011). When variability matters more than meaning: The effect of lexical forms on use of phonemic contrasts. *Developmental Psychology*, 47, 1448–1458.
- Thiessen, E. D., & Pavlik, P. I. Jr., (2013). IMinerva: A mathematical model of distributional statistical learning. *Cognitive Science*, 37, 310–343.
- Thiessen, E. D., & Yee, M. N. (2010). Dogs, bogs, labs, and lads: What phonemic generalizations indicate about the nature of children's early word-form representations. *Child Development*, 81(4), 1287–1303.
- Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, 104, 13273–13278.
- Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, 1, 197–234.
- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, 3, 1–30.
- Werker, J. F., & Tees, R. C. (1984). Cross language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63.
- White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, 59, 114–132.
- Winkelman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science*, 17, 799–806.
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-Month-olds use distinct object cues to categorize speech information. *Cognition*, 113, 234–243.
- Yoshida, K., Fennell, C., Swingle, D., & Werker, J. F. (2009). 14-month-olds learn similar-sounding words. *Developmental Science*.