Continuous developmental change explains discontinuities in word learning

Abdellah Fourtassi, Sophie Regan, Michael C Frank

Continuous developmental change explains discontinuities in word learning

Abdellah Fourtassi[1], Sophie Regan[1], & Michael C. Frank[1]

[1] Department of Psychology, Stanford University

Author Note

'The experiment, sample size, exclusion criteria, and the model's main predictions were preregistered at https://osf.io/942gv/'

'All data and analytic code are available at https://github.com/afourtassi/kidswitch'

'None of the authors have any financial interest or a conflict of interest regarding this work and this submission.'

'Acknowledgement: This work was supported by a post-doctoral grant from the Fyssen Foundation for AF, a summer internship funding by Stanford VPUE for SR, and Jacobs Foundation Research Fellowship program for MCF.'

Correspondence concerning this article should be addressed to Abdellah Fourtassi, Postal address. E-mail: abdellah.fourtassi@univ-amu.fr

Abstract

Cognitive development is often characterized in terms of discontinuities, but these discontinuities can sometimes be apparent rather than actual and can arise from continuous developmental change. To explore this idea, we use as a case study the finding by Stager and Werker (1997) that children's early ability to distinguish similar sounds does not automatically translate into word learning skills. Early explanations proposed that children may not be able to encode subtle phonetic contrasts when learning novel word meanings, thus suggesting a discontinuous/stage-like pattern of development. However, later work has revealed (e.g., through using more precise testing methods) that children do encode such contrasts, thus favoring a continuous pattern of development. Here we propose a probabilistic model that represents word knowledge in a graded fashion and characterizes developmental change as improvement in the precision of this graded knowledge. Our model explained previous findings in the literature and provided a new prediction — the referents' visual similarity modulates word learning accuracy. The models' predictions were corroborated by human data collected from both preschool children and adults. The broader impact of this work is to show that computational models, such as ours, can help us explore the extent to which episodes of cognitive development that are typically thought of as discontinuities may emerge from simpler, continuous mechanisms.

Keywords: word learning, cognitive development, computational modeling

Continuous developmental change explains discontinuities in word learning

Research Highlights

We provided a computational model of the development of word-pair learning, relating perceptual discrimination of pairs of sounds and objects to the ability to create distinct form-meaning associations.

The model characterizes this development in terms of a continuous process operating over similar representations across the lifespan.

The model makes novel predictions extending the work of Stager and Werker (1997). We successfully tested these predictions with both children and adults.

Introduction

Cognitive development is often characterized in terms of a succession of discontinuous stages. In Piaget's initial conception, these stages cross-cut different aspects of cognition (Piaget, 1954); in more modern conceptions, distinct domains are often thought to progress on their own timeline (e.g., Carey, Zaitchik, & Bascandziev, 2015). Although intuitively appealing, this sort of stage theory can be challenging to integrate with theories of learning, which typically posit that knowledge and skills improve incrementally with experience. Indeed, one of the central challenges of cognitive development has been to explain transitions between stages which appear to be qualitatively different (Carey, 2009).

Nevertheless, at least in some cases, development may only appear to be stage-like. Some discontinuities may be related to how we measure a specific skill. Other discontinuities may emerge due to statistical thresholding (e.g., an experimental p-value of $p < .05$ for one age group but not another) which can create a spurious dichotomy between success and failure in observing a given behavior. In such cases, positing discontinuous stages is unnecessary. Instead, a continuous model — involving similar representations across the

lifespan — may provide a simpler and more transparent account of development (cf. McMurray, 2007; Shultz, Schmidt, Buckingham, & Mareschal, 1995).

To explore this point computationally, we use a case study from word learning literature. Stager & Werker (1997) first showed that children's early ability to distinguish similar sounds does not automatically translate into word learning skills. The authors measured word learning using an audio-visual habituation Switch task. First, infants are familiarized with two word-object pairings (e.g., label 1 with object 1 and label 2 with object 2). Second, they are tested using two types of trials. The control "same" trial consists of a correct pairing (e.g., label 1 with object 1) and the "switch" trial consists of a wrong pairing (e.g., label 1 with object 2). If babies have correctly learned the association during the habituation, they are supposed to be surprised by the "switch" trial and not by the "same" trial. The former should thus result in a greater looking time compared to the latter (Werker, Cohen, Lloyd, Casasola, & Stager, 1998).

Though infants around 14-month old can distinguish perceptually similar sound pairs such as "dih" and "bih", they appear to fail in mapping this pair to two different objects in the switch task. This failure was initially taken as evidence that 14-month olds do not encode subtle sounds during meaning learning (Pater, Stager, & Werker, 2004; Stager & Werker, 1997). This interpretation suggested a discontinuous/stage-like pattern of development whereby younger children fail to encode the contrastive phonetic detail, whereas older children, around 17 months, typically do (Werker, Fennell, Corcoran, & Stager, 2002).

The initial discontinuous interpretation has been challenged by subsequent work. For instance, Yoshida, Fennell, Swingley, & Werker (2009) investigated whether failure in the Switch task reflects a lack of sound encoding during *habituation*, or whether it is only due to the nature of the *testing* method which does not allow learning below a certain threshold to be detected. They used the same habituation procedure as Stager & Werker (1997), but instead of comparing the looking times in "same" and "switch" trials, they tested infants

using a two-alternative choice task comparing fixations to target and distractor objects (Fernald, Perfors, & Marchman, 2006; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987). Using this testing method, researchers found evidence for learning even in 14-month olds.

Another challenge to the discontinuous account of development came from adult studies. If the mismatch between sound discrimination and word learning is only a stage in early infancy, then this mismatch should disappear by adulthood. Nonetheless, even adults show patterns of learning that mirror those shown by 14-month-olds when the sound contrasts are more challenging (Pajak, Creel, & Levy, 2016; White, Yee, Blumstein, & Morgan, 2013).

Some researchers (Pajak et al., 2016; Swingley, 2007; Yoshida et al., 2009) proposed that word knowledge may not be encoded in a binary fashion, i.e., it is not the case that children either succeed or fail in encoding minimal contrast when learning the meanings. Rather, they may be encoding this knowledge in a graded fashion (see Munakata (2001) for a detailed discussion of a similar view). Thus, development does not so much involve a qualitative shift (i.e., a sudden emergence of an ability that did not exist before) as much as it consists in the continuous refinement of initially noisy knowledge.

Many different computational formalisms can represent graded knowledge. Here we use probabilistic models, a formalism that allows both easy examinations of internal representations and quantification of the robustness of these representations. Our goal here is an exploration at the highest level of abstraction, in which we attempt to characterize the computational constraints that might lead to a particular behavior (Marr, 1982). We characterize word knowledge by a probability distribution in a similarity space. The probability is highest at the most typical sound instance of the word. It decreases as the instance becomes less typical. The precision of word knowledge can be characterized by whether it tolerates slightly atypical pronunciations. This tolerance is captured formally by the variance of the probability distribution: larger variance indicates higher tolerance and

lower precision, whereas smaller variance indicates lower tolerance and higher precision (for an illustration, see Figure 1 top and right panels).

This general framework — in which the precision of word knowledge is characterized with the variance of a probability distribution — can already provide an intuitive way of thinking about several findings. In particular, unlike the binary view, the probabilistic view allows for the possibility of word knowledge being both successful and noisy. This new understanding can provide an account for the fact that children show evidence of learning in some testing conditions (e.g., Yoshida et al., 2009) but not in others (e.g., Stager & Werker, 1997) — depending on the precision of the measurement.

In a word-pair learning paradigm, children are supposed to associate one label, e.g., "bih", with object 1 and a second label, e.g., "dih", with object 2. Infants may succeed in learning both associations. Nevertheless, the variance with which the pair of words are encoded can still be large, causing their probability distributions to overlap (Figure 1, top). The way this (noisy) knowledge is probed can lead to different results.

In the Switch task (Stager & Werker, 1997), children are understood to succeed if they reject a wrong association (e.g., "bih" with object 2). However, a large overlap between "bih" and "dih" means that "bih" is itself a plausible mispronunciation of "dih". The wrong association may not be rejected by children because the speaker could have said "bih" but meant "dih". In the two-alternative choice task (Yoshida et al., 2009), children do not have to reject the wrong association; they only need to show a preference, albeit small, for the correct one. Thus, unlike the Switch, this testing method allows us to see subtle evidence of learning even with a large overlap. For example, given the label "bih", children are supposed to pick which object is a better match to this label. Though it is possible that the speaker said "bih" and meant "dih", it is *more* likely that the speaker both said and meant "bih" — this higher probability leads to a preference for the correct object.

In addition to providing a potential explanation of the difference in behavior across the Switch and the preferential looking tasks, the probabilistic account explains differences in behavior within the same task. In particular, when the labels are quite distinct in the perceptual space ("lif" vs. "neem"), the probabilistic distributions do not overlap as much as in the case of similar-sounding words (Figure 1, left). This fact means that the learners will have less tolerance for the wrong association, leading to a successful rejection in the Switch task (as was reported by Stager & Werker (1997) and subsequent studies using the same paradigm). Further, distinctiveness can be enhanced even for minimally different sounds when other cues highlight their difference (Dautriche, Swingley, & Christophe, 2015; Rost & McMurray, 2009, 2010; Thiessen, 2007; Yeung & Werker, 2009).

In this framework, developmental change can be understood as an increase in the precision (i.e., a decrease in the variance) of the probabilistic knowledge, leading to a lower overlap between the distributions of similar-sounding words (Figure 1, right). Importantly, a more precise representation still has a non-zero variance. Thus, learning difficulties can still be induced with challenging stimuli or in cognitively demanding situations, as has been demonstrated in adult studies (Pajak et al., 2016; White et al., 2013).

The current study

A probabilistic account has been described before to explain patterns of learning and development at the qualitative level (e.g., Swingley, 2007). However, it is crucial to have a precise computational instantiation of this account which can help us 1) test this theoretical hypothesis more directly and 2) identify the particular parameters that are the locus of developmental change. One previous study attempted to provide such a computational instantiation (Hofer & Levy, 2017). However, this previous work aimed more at reproducing the results of a specific study (Pajak et al., 2016), which focused on explaining the mismatch between speech perception and word learning in adults rather than on exploring the mechanism of development.

The present work proposes a model of developmental change in word-pair learning based on the probabilistic account. We tested the ability of this model to both *explain* various findings in previous experiments in both children and adults (e.g., the fact that similar words are harder to learn than different words) and to *predict* new learning patterns that have not been tested before. In particular, we test the prediction that referent similarity (i.e., the confusability of pictures referred to by novel words) should play an identical computational role to word form similarity in predicting word recognition difficulty. Previous work in this line of research has studied the effect of several bottom-up and top-down properties on improving learning (e.g., Fennell & Waxman, 2010; Rost & McMurray, 2009; Thiessen, 2007) but the effect of the referent similarity remains to be tested. While some researchers studied this effect on wordform learning (Fourtassi & Dupoux, 2019; Yeung & Werker, 2009; Zhao, Packard, McMurray, & Gupta, 2019), to our knowledge, the effect on children's word-object mapping has not been investigated before. Finally, we explore the extent to which the probabilistic account allows us to understand development in terms of as a continuous refinement in similar representations across the lifespan.

The paper is organized as follows. First, we introduce the model and we explain how it allows us to characterize behavior in a word-pair-learning paradigm. Then we explore the predictions of the model through simulating its behavior across different parameter settings. Next, we quantify the extent to which the model's predictions account for human data we collected from both preschool children and adults. Finally, we discuss the results in light of existing accounts of word development.

Model

Probabilistic structure

Our model consists of a set of variables describing the general process of spoken word recognition in a referential situation. These variables are related in a way that reflects the simple generative scenario represented graphically in Figure 2. When a speaker utters a

sound in the presence of an object, the observer assumes that the object o activated the concept C in the speaker's mind. The concept prompted the corresponding label L. Finally, the label was physically instantiated by the sound s.

A similar probabilistic structure was used by Lewis & Frank (2013) to model concept learning, and by Hofer & Levy (2017) to model spoken word learning. However, the first study assumed that the sounds are heard unambiguously, and the second assumed the concepts are observed unambiguously. In our model, we make the — more realistic — assumption that both labels and concepts are characterized with distributions and that the distributions of neighboring categories may overlap. In other words, we allow the model to account for the possibility that an instance of a given category (e.g.,"bih") can be mistaken for an instance of a neighboring category (e.g., "dih"), albeit with a smaller probability. We assume, for simplicity, that this probability is captured by a normal distribution[1]:

$$p(o|C) \sim \mathcal{N}(\mu_C, \sigma_C^2) \tag{1}$$

and

$$p(s|L) \sim \mathcal{N}(\mu_L, \sigma_L^2) \tag{2}$$

Finally, we assume there to be one-to-one mappings between concepts and labels and

[1] We used a normal distribution to simplify the mathematical derivations. This is obviously not aimed at providing a precise representation for the word-forms (and even less so for the visual objects), but as we shall see in the results section, this simplified representation was enough to quantify the confusability of a given stimulus with its neighbors, which is the quantity we needed to predict the accuracy of word-object mapping in our task.

that observers have successfully learned these mappings during the exposure phase:

$$P(L_i|C_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Inference

In our canonical inference case, the learner hears a sound s and has to decide which object o provides an optimal match to this sound (see Figure 3). To this end, they must compute the probability $P(o|s)$ for all possible objects. This probability can be computed by summing over all possible concepts and labels:

$$P(o|s) = \sum_{C,L} P(o, C, L|s) \tag{4}$$

Using the fact that $P(o, C, L|s) = \frac{P(o,C,L,s)}{P(s)}$ and that $P(s)$ does not depend on o, we arrive at the equation:

$$P(o|s) \propto \sum_{C,L} P(o, C, L, s) \tag{5}$$

The joint probability $P(o, C, L, s)$ is obtained by factoring the graphical model in Figure 2:

$$P(o, C, L, s) = P(s|L)P(L|C)P(C|o)P(o)$$

Using Bayes' rule, we can rewrite $P(C|o)$ in terms of $P(o|C)$:

$$P(C|o) = \frac{P(o|C)P(C)}{P(o)}$$

By subtituting this term in the expression of the joint distribution $P(o, C, L, S)$ we obtain:

$$P(o, C, L, s) = P(s|L)P(L|C)P(o|C)P(C)$$

Finally, assuming that the concepts' prior probability $P(C)$ is uniformly distributed,[2] we obtain the following expression, where all conditional dependencies are now well defined:

$$P(o|s) \propto \sum_{C,L} P(s|L)P(L|C)P(o|C) \tag{6}$$

Task and model predictions

We use the model to predict word learning in a task similar to the one introduced by Stager & Werker (1997). We used a modified version of the task where the testing method consists in a two-alternative forced-choice (Yoshida et al., 2009). In this task, participants are first exposed to two different word-object pairings (e.g., "lif" - object 1, "neem" - object 2). The word-object associations are introduced sequentially. After this exposure phase, participants perform a series of test trials. In each of these trials, one of the two sounds is uttered (e.g., "lif") and participants choose the corresponding object from the two alternatives. An overview of the task is shown in Figure 3.

From the general expression 6, we derive three exact analytical solutions instantiating different learning assumptions (see the Appendix for the full mathematical derivation).

Model 1. Recall from expressions 1 and 2 that $P(o|C)$ and $P(s|L)$ have parameters σ_C and σ_L, respectively, that control for uncertainty. The first solution/model is derived by assuming that the labels are recovered from sounds with a certain level of uncertainty

[2] This is a reasonable assumption in our particular case given the similarity of the concept pairs used in each naming situation in the experiment.

$\sigma_L > 0$, but that concepts are unambiguously recovered from the observed objects, i.e., $\sigma_C \to 0$. This assumption has been made — whether implicitly or explicitly — by most previous work in this line of research. For example, in Stager & Werker (1997), the objects were quite dissimilar. Thus, the assumption that they were easily discriminated by infants seems relatively well justified. One important implication of this assumption is that only the similarity of word sounds modulates performance in word learning, not the similarity of the referents (as long as these referents are differentiated perceptually).

Using this assumption, we can simplify the general Equation 6 of the probability $P(o_i|s_i)$ of choosing the correct object o_i given the correspondding label instance s_i:

$$P(o_i|s_i) = \frac{1}{1 + e^{-\frac{\Delta s^2}{2\sigma_L^2}}} \tag{7}$$

where $\Delta s = s_i - s_j$ is the perceptual sound distance between the pair of sounds that we refer to as s_i and s_j.

Model 2. The second solution/model is derived from Equation 6 by making the more general assumption that both the labels and the concepts are recovered with uncertainty from the sounds and objects, but we introduce the simplifying assumption that the label-related uncertainty σ_L and the concept-related uncertainty σ_C are of a similar magnitude, i.e., $\sigma_C \approx \sigma_L = \sigma$. This simplifying assumption makes the prediction that the sound similarity and the object similarity impact word learning accuracy in exactly the same way. Furthermore, it allows us to study the behavior of the model with only one free parameter, an important consideration given the small number of datapoints available from any given developmental experiment.

$$P(o_i|s_i) = \frac{1 + e^{-\frac{\Delta s^2 + \Delta o^2}{2\sigma^2}}}{1 + e^{-\frac{\Delta s^2 + \Delta o^2}{2\sigma^2}} + e^{-\frac{\Delta s^2}{2\sigma^2}} + e^{-\frac{\Delta o^2}{2\sigma^2}}} \tag{8}$$

Model 3. We finally derive the third (and most general) solution which allows label- and concept-related uncertainties to vary independently.

$$P(o_i|s_i) = \frac{1 + e^{-(\frac{\Delta s^2}{2\sigma_L^2} + \frac{\Delta o^2}{2\sigma_C^2})}}{1 + e^{-(\frac{\Delta s^2}{2\sigma_L^2} + \frac{\Delta o^2}{2\sigma_C^2})} + e^{-\frac{\Delta s^2}{2\sigma_L^2}} + e^{-\frac{\Delta o^2}{2\sigma_C^2}}} \tag{9}$$

In order to understand the predictions of the models (especially the more general ones, i.e., Model 2 and 3), Figure 4 show simulations of the accuracy $P(o_i|s_i)$ as a function of the distinctiveness parameters (Δs and Δo) and the uncertainty parameters σ_L and σ_C.

The simulations explain two experimental results from previous studies and make one new prediction:

1) For fixed values of Δo and σ, the probability of accurate responses increases as a function of Δs. This pattern accounts for the fact that similar sounds are generally more challenging to learn than different sounds for both children (Stager & Werker, 1997) and adults (Pajak et al., 2016).

2) For fixed values of Δs and Δo, accuracy increases when the representational uncertainty σ decreases. This observation provides a simple model for developmental change. Younger children have noisier representations (see Swingley, 2007; Yoshida et al., 2009), which leads to lower word recognition accuracy, especially for similar-sounding words.

3) For fixed values of Δs and σ, accuracy increases with the visual distance between the semantic referents Δo.

Experiment

In this experiment, we tested participants in the word learning task introduced above (Figure 3). More precisely, we explored the predictions related to both distinctiveness and

precision. Sound similarity (Δs) and object similarity (Δo) were varied simultaneously in a within-subject design. Two age groups (preschool children and adults) were tested on the same task[3] to explore whether development can be characterized with the uncertainty parameters, σ_C and σ_L. The experiment, sample size, exclusion criteria, and the model's main predictions were pre-registered.[4]

Methods

Participants. We report data from $N = 63$ children ages 4-5 years from a nursery school. An additional $N = 39$ children participated but were removed from analyses using a preregistered exclusion criterion of not being above chance on the catch trials.[5] We also report data from $N = 74$ adult participants tested on Amazon Mechanical Turk. An additional $N = 26$ were tested but removed from analyses using preregistered exclusion criteria of either having a low score on the catch trials or being familiar with the non-English sound stimuli we used in the adult experiment.

Stimuli and similarity rating. The sound stimuli were generated using the MBROLA Speech Synthesizer (Dutoit, Pagel, Pierret, Bataille, & Van der Vrecken, 1996). We generated three kinds of nonsense word pairs which varied in their degree of perceptual similarity to English speakers: 1) *different* pairs: "lif"/"neem" and "zem"/"doof", 2) *intermediate* pairs: "aka"/"ama" and "ada"/"aba", and 3) *similar* non-English pairs[6]:

[3] This four-condition within-subject design is somewhat unusual for preschoolers, but we followed the tablet paradigm (e.g., Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016) which allowed us to gather a relatively large number of trials from each child.

[4] See repository at https://github.com/afourtassi/kidswitch.

[5] The number of children who were excluded is a bit high for this age group. This is partly due to our adoption of a 2x2 fully within-subject design, which was demanding for children. As we explain below, we adopted this design based on pilot testing that indicated that precision of measurement was critical for testing our experimental predictions.

[6] For the intermediate/similar pairs, we made the choice of using a VCV (Vowel-Consonant-Vowel) structure in order to emphasize the phonemic difference in the minimal pairs, especially for the non-native contrasts.

"ada"/"ad^{h}a" (in hindi) and "aʕa"/"aħa" (in arabic). For the similar pairs, we used non-native sounds because this was a good way to decrease perceptual similarity for adults. In this respect, we took inspiration from the work of Pajak et al. (2016). We refer the reader to that paper for a lengthy discussion of the literature on (and implications for) L2 phoneme perception.

As for the objects, we used the Dynamic Stimuli javascript library[7] which allowed us to generate objects in four different categories: "tree," "bird," "bug," and "fish." These categories were described to participants as naturally occurring kinds on an alien planet. In each category, we generated *different*, *intermediate*, and *similar* pairs by manipulating a continuous property controlling features of the category's shape (e.g., body stretch or head fatness).

In order to validate and quantify our similarity scales, we ran a separate survey on Amazon Mechanical Turk where we asked $N = 20$ adult participants to evaluate the similarity of each sound and object pair on a 7-point scale. Data are shown in Figure 5 where we scaled responses within the range [0,1] for each stimulus group. We used these data in all models as an empirical measurement of the perceptual distance between the sound pairs and the object pairs. The use of empirical measurement allows us to eliminate Δs and Δs as free parameters (see Frank & Goodman (2012) and Xu & Tenenbaum (2007) for a similar strategy).

Design. Each age group saw only two of the three levels of similarity described in the previous sub-section: *different* vs. *intermediate* for the preschoolers, and *intermediate* vs.

However, as one reviwer rightfully pointed out to us, this structure could influence word leaning performance since it is not common in English. Nevertheless, if we observed this effect, we would have seen differences in predictability between the intermediate/similar pairs on the one hand, and different pairs (which are phonotactically common in English) on the other hand. However, this was not the case; word learning accuracy was correctly predicted based on perceptual similarity alone.

[7] https://github.com/erindb/stimuli

similar for adults. We made this choice in light of pilot studies showing that adults were at ceiling with *different* sounds/objects, and children were at chance with the *similar* sounds/objects. That said, this difference in the level of similarity is accounted for in the model: We used empirical distance measurement to fill in the appropriate values of Δs and Δo for each age group.

To maximize our ability to measure subtle stimulus effects, the experiment was a 2x2 within-subjects factorial design with four conditions: high/low sound similarity crossed with high/low visual object similarity. Besides the four conditions, we also tested participants on a fifth catch condition which was similar in its structure to the other ones but was trivially easy and used only to select participants who were able to follow the instructions and show minimal learning.

Procedure. Preschoolers were tested at the nursery school using a tablet, whereas adults used their own computers to complete the same experiment online. Participants were tested in a random sequence of five conditions: the four experimental conditions plus the catch condition. In each condition, participants saw a first block of four exposure trials followed by four testing trials, and a second block of two exposure trials (for memory refreshment) followed by an additional four testing trials. The length of this procedure was demanding, especially for children, but we adopted a fully within-subjects design based on pilot testing that indicated that precision of measurement was critical for testing our experimental predictions.

In the exposure trials, participants saw two objects associated with their corresponding sounds. We presented the first object on the left side of the tablet's screen simultaneously with the corresponding sound. The second sound-object association followed on the other side of the screen after 500ms. For both objects, visual stimuli were present for the duration of the sound clip (about 800ms). During this phase, participants were asked to pay attention and try yo lean the words. In the testing trials, participants saw both objects simultaneously

and heard only one sound. They completed the trial by selecting which of the two objects corresponded to the sound. The object-sound pairings were randomized across participants, as was the order of the conditions (except for the catch condition which was always placed in the middle of the testing sequence). We also randomized the on-screen position (left vs. right) of the two pictures on each testing trial.

Results

Experimental results are shown in Figure 6 (solid lines). We first analyzed the results using a mixed-effects logistic regression with sound distance, object distance and age group as fixed effects, and with a maximal random effects structure (allowing us to take into account the full nested structure of our data) (Barr, Levy, Scheepers, & Tily, 2013). We found main effects for all the fixed effects in the regression. For the sound distance, we obtained $\beta = 0.68$ ($p < 0.001$), replicating previous findings that sound distance modulates success in word learning (e.g., Stager & Werker, 1997).

For object distance, we found $\beta = 0.60$ ($p < 0.001$), and this finding confirms the new prediction of our model, according to which, object distance also modulates performance in word learning. Finally, for the age group, we obtained $\beta = 0.59$ ($p < 0.001$), showing that overall performance improves with age. The full output of the regression model is shown in Table 2.

We next fit the three models obtained through expressions 7, 8, and 9 to the population's mean responses in each age group, using a Nonlinear Least Squares algorithm. The predictions of the models are shown 6. The parameter estimates (for σ_L and σ_C) as well as models' goodness to fit (i.e., measured through R^2) are presented in Table 1.

Model 1, which does not take into account ambiguity in recovering concepts from observed objects, explains only a small part of the variance. In contrast, Model 3, which does take into account this ambiguity, accounts for all the variance. Interestingly, Model 2

which has a single, shared uncertainty parameter for both auditory and visual modalities still explains almost all the variance in human data.

As predicted, the uncertainty parameters were larger for children than they were for adults (Table 1), showing that word knowledge gets more precise with development. Further, the parameter estimates of Model 3 show that this developmental effect is larger for labels (σ_L varies between 0.82 in children and 0.12 in adults) than it is for concepts (σ_C varies between 0.33 in children and 0.17 in adults).

General Discussion

This paper explored the idea that some seemingly stage-like patterns in cognitive development can be characterized in a continuous fashion. We used as a case study the seminal work of Stager & Werker (1997) showing a discrepancy between children's speech perception abilities and their word learning skills. The development of this discrepancy could be understood in terms of a discrete change in word representation. But our model demonstrates that it can also be parsimoniously described as a result of continuous developmental change in the precision of children's graded word knowledge. Our model thus provides a computational instantiation and test of the continuous development hypothesis proposed in previous work (Pajak et al., 2016; Swingley, 2007; Yoshida et al., 2009).

Our model used the assumption that the ability to distinguish two sounds perceptually leads to an accurate mapping of these sounds with two distinct objects, although this mapping can be noisy. On this account, apparent patterns of failure and success in retrieval tasks (such as ours) reflect different degrees of noise in the learned representations and not necessarily the presence/absence of these representations. With this assumption, the model was capable of accurately predicting behavior (including around-chance performance in the case of the most difficult contrasts)[8], thus providing a proof of concept for the continuity

[8] That is, the case when both the sounds and objects are similar for children. Figure 6 shows that the

hypothesis. The discontinuity hypothesis — which posits the absence of learning and not just retrieval difficulty — cannot be ruled out by our model, but it is less parsimonious. We propose that the continuous hypothesis should be considered as a baseline; discontinuous development should only be posited if it deviates from this baseline.

We find in the literature two broad accounts of development in the ability to learn the meanings of similar-sounding words: One that suggests *direct* development of the sound representation and one that hypothesizes *indirect* development of this representation through improvement in general cognitive resources. On the first account, the sound representation becomes more precise as learners refine the boundaries of their initially ambiguous phonetic categories and as they gain more experience with the functional role of these categories (Apfelbaum & McMurray, 2011; Dietrich, Swingley, & Werker, 2007; Rost & McMurray, 2009, 2010; Werker & Curtin, 2005; Yoshida et al., 2009). On the second account, the precision of sound encoding in the switch task improves as a result of the maturation of more general resources like the attentional and working memory capacity (Hofer & Levy, 2017; Stager & Werker, 1997; Werker & Fennell, 2004). Such improvement allows older children and adults to better encode the sound details while simultaneously matching these sounds to visual objects. Indeed, one recent meta-analysis of the switch task concluded that both changing representation precision and better memory/attention play a role in developmental changes (Tsui, Byers-Heinlein, & Fennell, 2019).

Our model is compatible with both of these accounts. Indeed, the model aims at understanding *how* uncertainty can influence word learning. It characterizes and measures uncertainty in a rather global/coarse fashion, without committing to any specific source. Thus, it is compatible[9] with direct accounts that posit uncertainty primarily stems from

---

confidence intervals include the chance value of 0.5.

[9] The fact that our model is compatible with several previous accounts does not necessarily mean that these accounts are themselves compatible with one another.

noisy phonological representations, and with the indirect account which suggests uncertainty primarily stems from a limitation in the general cognitive resources, hindering precise encoding in memory of subtle sounds under challenging conditions. Note that, since our goal was not to distinguish between these accounts, the model (at least in its current format) cannot answer questions about the development of each of these sources of uncertainty separately or about their relative contribution to the global uncertainty.

While most research has focused on sound representation specifically in analyzing the process of learning similar-sounding words, this work showed that the visual representation of the referent is equally important. Indeed, Model 1 — which assumes that any visually discriminable contrast can be encoded unambiguously as separate referents — failed to explain the data, whereas Model 2 and 3 — which take into account visual ambiguity — succeeded. As a consequence of this assumption, we found that just like word learning is modulated by the phonological similarity of the form, it is also modulated by the visual similarity of the semantic referents. This result is different from previous findings on the effect of the referents' similarity on the phonological development of the wordform (Fourtassi & Dupoux, 2019; Yeung & Werker, 2009; Zhao et al., 2019). Our result is about wordform-referent mapping — a different learning problem as it involves, in particular, the encoding of visual semantic/referent knowledge about the world (see Zhao et al. (2019) for a detailed discussion). While we did not model the process of phonological development per se, we argued in our discussion above that part of the observed difference between children and adults in terms of their word-referent mapping skills can be attributed to phonological development.

Model 2, which predicts that sound similarity and visual similarity influence word learning accuracy in the same way, explained slightly less variance than Model 3 which predicts that these modalities influence word learning differently. Further, as we stated in the results section, a comparison of the variance estimates across age groups showed that

uncertainty reduction in the visual modality was lower compared to that of the auditory modality (Table 1). Perhaps this difference is due to the fact that, in our task, the auditory speech had more sources of noise — that children have to deal with — than the visual input did. The processing of speech involved dealing with both perceptual noise and categorical ambiguity (due to the fact that the phonemic boundaries are still developing). In contrast, the processing of the visual input in our task involved only perceptual noise and no category-related uncertainty. A future direction of research is independent measurement and comparison of these parameters in children.

Our finding that word learning is mediated by the visual similarity of the semantic objects has implications for theories of lexical development. It suggests that, all things being equal, children may learn, first, words whose semantic referents are visually different as this allows them to minimize semantic ambiguity. It will be interesting for future work to explore whether the results that we obtained using visual similarity generalize to richer, more conceptual features in the semantic space. In addition, it is important to study how laboratory experiments of this sort may explain patterns of word learning in the wild (Engelthaler & Hills, 2017; Fourtassi, Bian, & Frank, 2018; Sizemore, Karuza, Giusti, & Bassett, 2018).

Our work found, among other things, that children have higher sound/phoneme uncertainty than adults do, but prior work have shown that children have good auditory encoding (Swingley, 2003; Swingley & Aslin, 2000, 2002). These two findings are not in conflict with each other. In this work, and as we explained above, the measured sound uncertainty includes various sources. One of these sources is perceptual uncertainty, i.e., the ability to perceptually distinguish a sound from other neighboring sounds (e.g., “bin” is perceptually different from “din”), and another source is categorical uncertainty, i.e., the ability to estimate phonological boundaries, e.g., “bin” and “din” straddle a phonological boundary, that is, they signal a change in meaning. The fact that children have a higher

sound uncertainty than adults do, does not necessarily mean this higher uncertainty is due to poorer perceptual encoding only. In fact, it is possible that much of the differential in uncertainty (compared to adults) is due to the categorical/phonological dimension which we know undergoes a process of reorganization and refinement across development. Thus, our finding that children have higher sound-related uncertainty does not contradict previous work showing children are capable of fine-grained speech perception (see Werker & Curtin (2005) for a similar view).

There are a number of limitations to this work. One is that the model was fit to data from children at a relatively older age (4-5 years old) than what is typically studied in the literature (14-17 month-old). We selected this older age group to optimize the number and precision of the experimental measures (both are crucial to model fitting). Data collection involved presenting participants with several trials across four conditions in a within-subject design. It would have been challenging to obtain such measures with infants. That said, though we used data from older children, we still found clear developmental differences with adults, confirming and extending findings that the ability to distinguish similar-sounding words continues developing well beyond 17 months (Fennell & Byers-Heinlein, 2014; Hazan & Barrett, 2000; Mattock, Polka, Rvachew, & Krehm, 2010).

One limitation of our models is that they only account for bottom-up, similarity-based effects. They do not account for how high-level factors such as social and communicative cues can influence learning. For example, Fennell & Waxman (2010) highlighted the fact that some laboratory tasks such as the one used in Stager & Werker (1997) introduce novel words in isolation (e.g., “neem!”) rather than within a naming phrase (e.g., “look at the neem!”). This fact may prompt children to interpret these novel words in a non-referential way (e.g., an exclamation such as “Wow!”).

To conclude, this paper proposes a model that accounts for the development of an important aspect of word learning. Our account suggests that the developmental data can be

explained based on a continuous process operating over similar representations across the lifespan, suggesting developmental continuity. We used a case from word learning as an example, but the same idea might apply to other aspects of cognitive development that are typically thought of as stage-like (e.g., acquisition of a theory of mind). Computational models, such as the one proposed here, can help us investigate the extent to which such discontinuities emerge due to genuine qualitative changes and the extent to which they reflect the granularity of the researchers' own measurement tools.

Appendix

Here we provide the full mathematical derivation of the models' predictions. More specifically, we explain how the general Equation 6 leads to Equations 7, 8, and 9 which represent the three models. Each of these models instantiates an approximation of $P(o_i|s_i)$; the probability that, given a label instance s_i, a participant would choose the correct object o_i, as opposed to the incorrect object o_j, where $i, j \in \{1, 2\}$ and $j = 1 - i$.

We start with the general expression of this probability as defined in Equation 6. This expression can be normalized in our special case (i.e., a two-alternative choice task) as follows:

$$P(o_i|s_i) = \frac{\sum_{C,L} P(s_i|L)P(L|C)P(o_i|C)}{\sum_{C,L} P(s_i|L)P(L|C)P(o_i|C) + \sum_{C,L} P(s_i|L)P(L|C)P(o_j|C)}.$$

Using the definition of $P(L|C)$ (given in Equation 3) and the fact that there are two categories in each modality (L_1, L_2 and C_1, C_2) we obtain:

$$P(o_i|s_i) = \frac{P(s_i|L_i)P(o_i|C_i) + P(s_i|L_j)P(o_i|C_j)}{[P(s_i|L_i)P(o_i|C_i) + P(s_i|L_j)P(o_i|C_j)] + [P(s_i|L_i)P(o_j|C_i) + P(s_i|L_j)P(o_j|C_j)]} \tag{10}$$

In what follows, we compute the explicit expression of each conditional probability in Equation 10. $P(s_i|L_i)$ is the probability that the sound s_i was sampled from its correct category L_i. This probability follows a normal distribution $\mathcal{N}(\mu_C, \sigma_C^2)$ (Equation 2), thus:

$$P(s_i|L_i) = \frac{1}{\sqrt{2\pi\sigma_{L_i}^2}} e^{-\frac{(s_i-\mu_{L_j})^2}{2\sigma_{L_i}^2}}.$$

We use the simplifying assumption that $\sigma_{L_i} = \sigma_L$, an uncertainty common to all labels. σ_L can be understood as characterizing general speech processing skills (at a given time in development). We obtain the following expression:

$$P(s_i|L_i) = \frac{1}{\sqrt{2\pi\sigma_L^2}} e^{-\frac{(s_i-\mu_{L_j})^2}{2\sigma_L^2}}.$$

s_i is supposed to be the prototypical instance of L_i, that is, $s_i = \mu_{L_j}$, leading to

$$P(s_i|L_i) = \frac{1}{\sqrt{2\pi\sigma_L^2}}.$$

Similarly,

$$P(o_i|C_i) = P(o_j|C_j) = \frac{1}{\sqrt{2\pi\sigma_C^2}}.$$

$P(s_i|L_j)$ is the probability that the sound s_i, though a perfect instance of L_i, could still be a mispronounced instance of the category L_j. The value of this probability depends on the extent to which the distributions of the categories L_i and L_j overlap with each other (Figure 1). Using the normal distribution, we obtain the following expression:

$$P(s_i|L_j) = \frac{1}{\sqrt{2\pi\sigma_L^2}} e^{-\frac{(s_i-s_j)^2}{2\sigma_L^2}} = \frac{1}{\sqrt{2\pi\sigma_L^2}} e^{-\frac{\Delta s^2}{2\sigma_L^2}}.$$

Similarly,

$$P(o_i|C_j) = P(o_j|C_i) = \frac{1}{\sqrt{2\pi\sigma_C^2}} e^{-\frac{\Delta o^2}{2\sigma_C^2}}$$

Finally, we substitute all these expressions in equation 10 above, leading to:

$$P(o_i|s_i) = \frac{1 + e^{-(\frac{\Delta s^2}{2\sigma_L^2} + \frac{\Delta o^2}{2\sigma_C^2})}}{1 + e^{-(\frac{\Delta s^2}{2\sigma_L^2} + \frac{\Delta o^2}{2\sigma_C^2})} + e^{-\frac{\Delta s^2}{2\sigma_L^2}} + e^{-\frac{\Delta o^2}{2\sigma_C^2}}}$$

which is the expression for Model 3 (Equation 9). The expression for Model 2 (Equation 8) can be deduced from that of Model 3 by positing $\sigma = \sigma_L = \sigma_C$. The expression for Model 1 (Equation 7) is also derived from Model 3 by taking the limit when $\sigma_C \to 0$.

All data and code for these analyses are available at

https://github.com/afourtassi/kidswitch

References

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*.

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3).

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Carey, S., Zaitchik, D., & Bascandziev, I. (2015). Theories of development: In dialog with jean piaget. *Developmental Review*, *38*.

Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition*, *143*.

Dietrich, C., Swingley, D., & Werker, J. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences*, *104*.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). The mbrola project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *Proceedings of ICSLP* (Vol. 3). IEEE.

Engelthaler, T., & Hills, T. T. (2017). Feature biases in early word learning: Network distinctiveness predicts age of acquisition. *Cognitive Science*, *41*.

Fennell, C., & Byers-Heinlein, K. (2014). You sound like mommy: Bilingual and monolingual infants learn words best from speakers typical of their language environments. *International Journal of Behavioral Development*, *38*(4).

Fennell, C., & Waxman, S. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, *81*.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*.

Fourtassi, A., Bian, Y., & Frank, M. C. (2018). Word learning as network growth: A cross-linguistic analysis. In *Proceedings of the 40th annual meeting of the cognitive science society*.

Fourtassi, A., & Dupoux, E. (2019). Phoneme learning is influenced by the taxonomic organization of the semantic referents. In *Proceedings of the 41st annual meeting of the cognitive science society*.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084).

Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*,

17(1).

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*(4).

Hofer, M., & Levy, R. (2017). Modeling Sources of Uncertainty in Spoken Word Learning. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*.

Lewis, M., & Frank, M. C. (2013). An integrated model of concept learning and word-concept mapping. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35).

Marr, D. (1982). *Vision*. WH Freeman.

Mattock, K., Polka, L., Rvachew, S., & Krehm, M. (2010). The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science*, *13*(1).

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*(5838).

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences*, *5*(7).

Pajak, B., Creel, S., & Levy, R. (2016). Difficulty in learning similar-sounding words: A developmental stage or a general property of learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9).

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological

contrasts. *Language*, *80*.

Piaget, J. (1954). *The construction of reality in the child*. New York, NY, US: Basic Books.

Rost, G., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*.

Rost, G., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*.

Shultz, T. R., Schmidt, W. C., Buckingham, D., & Mareschal, D. (1995). Modeling cognitive development with a generative connectionist algorithm. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sizemore, A. E., Karuza, E. A., Giusti, C., & Bassett, D. S. (2018). Knowledge gaps in the early growth of semantic feature networks. *Nature Human Behaviour*, *2*(9).

Stager, C., & Werker, J. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640).

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech*, *46*(2-3), 265–294.

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology*, *43*(2).

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*(5), 480–484.

Thiessen, E. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, *56*.

Tsui, A. S. M., Byers-Heinlein, K., & Fennell, C. (2019). Associative word learning in infancy: A meta-analysis of the switch task. *Developmental Psychology*, *55*.

Werker, J., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, *34*.

Werker, J., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*.

Werker, J., & Fennell, C. (2004). Listening to sounds versus listening to words: Early steps in word learning. In D. G. Hall & S. Waxman (Eds.), *Weaving a lexicon*. Cambridge: MIT Press.

Werker, J., Fennell, C., Corcoran, K., & Stager, C. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*.

White, K., Yee, E., Blumstein, S., & Morgan, J. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language*, *68*(4).

Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological Review*, *114*(2).

Yeung, H., & Werker, J. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*.

Yoshida, K., Fennell, C., Swingley, D., & Werker, J. (2009). 14-month-olds learn similar-sounding words. *Developmental Science*, *12*.

Zhao, L., Packard, S., McMurray, B., & Gupta, P. (2019). Similarity of referents influences the learning of phonological word forms: Evidence from concurrent word learning. *Cognition*, *190*, 42–60.

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*.

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3).

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Carey, S., Zaitchik, D., & Bascandziev, I. (2015). Theories of development: In dialog with jean piaget. *Developmental Review*, *38*.

Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition*, *143*.

Dietrich, C., Swingley, D., & Werker, J. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences*, *104*.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). The mbrola project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *Proceedings of ICSLP* (Vol. 3). IEEE.

Engelthaler, T., & Hills, T. T. (2017). Feature biases in early word learning: Network distinctiveness predicts age of acquisition. *Cognitive Science*, *41*.

Fennell, C., & Byers-Heinlein, K. (2014). You sound like mommy: Bilingual and monolingual infants learn words best from speakers typical of their language environments.

International Journal of Behavioral Development, *38*(4).

Fennell, C., & Waxman, S. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, *81*.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*.

Fourtassi, A., Bian, Y., & Frank, M. C. (2018). Word learning as network growth: A cross-linguistic analysis. In *Proceedings of the 40th annual meeting of the cognitive science society*.

Fourtassi, A., & Dupoux, E. (2019). Phoneme learning is influenced by the taxonomic organization of the semantic referents. In *Proceedings of the 41st annual meeting of the cognitive science society*.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084).

Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*, *17*(1).

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*(4).

Hofer, M., & Levy, R. (2017). Modeling Sources of Uncertainty in Spoken Word Learning. In

Proceedings of the 39th Annual Meeting of the Cognitive Science Society.

Lewis, M., & Frank, M. C. (2013). An integrated model of concept learning and word-concept mapping. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35).

Marr, D. (1982). *Vision*. WH Freeman.

Mattock, K., Polka, L., Rvachew, S., & Krehm, M. (2010). The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science*, *13*(1).

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*(5838).

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences*, *5*(7).

Pajak, B., Creel, S., & Levy, R. (2016). Difficulty in learning similar-sounding words: A developmental stage or a general property of learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9).

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language*, *80*.

Piaget, J. (1954). *The construction of reality in the child*. New York, NY, US: Basic Books.

Rost, G., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*.

Rost, G., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*.

Shultz, T. R., Schmidt, W. C., Buckingham, D., & Mareschal, D. (1995). Modeling cognitive

development with a generative connectionist algorithm. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sizemore, A. E., Karuza, E. A., Giusti, C., & Bassett, D. S. (2018). Knowledge gaps in the early growth of semantic feature networks. *Nature Human Behaviour*, *2*(9).

Stager, C., & Werker, J. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640).

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech*, *46*(2-3), 265–294.

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology*, *43*(2).

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*(5), 480–484.

Thiessen, E. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, *56*.

Tsui, A. S. M., Byers-Heinlein, K., & Fennell, C. (2019). Associative word learning in infancy: A meta-analysis of the switch task. *Developmental Psychology*, *55*.

Werker, J., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, *34*.

Werker, J., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech

processing. *Language Learning and Development*, *1*.

Werker, J., & Fennell, C. (2004). Listening to sounds versus listening to words: Early steps in word learning. In D. G. Hall & S. Waxman (Eds.), *Weaving a lexicon*. Cambridge: MIT Press.

Werker, J., Fennell, C., Corcoran, K., & Stager, C. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*.

White, K., Yee, E., Blumstein, S., & Morgan, J. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language*, *68*(4).

Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological Review*, *114*(2).

Yeung, H., & Werker, J. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*.

Yoshida, K., Fennell, C., Swingley, D., & Werker, J. (2009). 14-month-olds learn similar-sounding words. *Developmental Science*, *12*.

Zhao, L., Packard, S., McMurray, B., & Gupta, P. (2019). Similarity of referents influences the learning of phonological word forms: Evidence from concurrent word learning. *Cognition*, *190*, 42–60.

Table 1

Characteristics and performance of the models used in this study. Value ranges represent 95% confidence intervals.

				Children		Adults	
Model	Structure	Param.	R^2	σ_L	σ_C	σ_L	σ_C
model 1	σ_L only	1	0.27	1.01 [0.55, 1.47]	–	0.37 [0.01, 0.73]	–
model 2	$\sigma_L = \sigma_C$	1	0.96	0.6 [0.53, 0.67]	0.6 [0.53, 0.67]	0.15 [0.12, 0.17]	0.15 [0.12, 0.17]
model 3	$\sigma_L \neq \sigma_C$	2	1.00	0.82 [0.69, 0.95]	0.33 [0.18, 0.48]	0.12 [0.12, 0.13]	0.17 [0.16, 0.18]

	Predictor	Estimate	Std.Error	z.value	p.value
1	(Intercept)	1.06	0.21	5.02	<0.01
2	sound_dist	0.68	0.14	4.72	<0.01
3	object_dist	0.6	0.15	3.96	<0.01
4	age	0.59	0.16	3.64	<0.01
5	sound_dist*object_dist	0.36	0.14	2.56	0.01
6	sound_dist*age	0.37	0.13	2.83	<0.01
7	object_dist*age	0.25	0.13	1.91	0.06
8	sound_dist*object_dist*age	0.19	0.13	1.45	0.15

Table 2

Predictor estimates with standard errors and significance information for a logistic mixed-effects model predicting the accuracy of word learning.

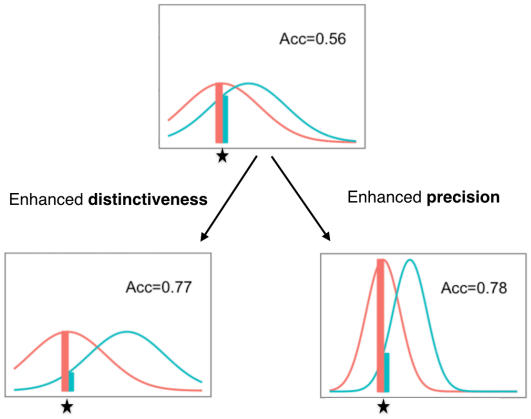

Figure 1. An illustration of the probabilistic/continuous account using simulated data. A word is represented with a distribution over the perceptual space (indicated in red or blue). When the uncertainty of the representation is large relative to the distance between the stimuli (top panel), an instance of the red category (indicated with a star) could also be a plausible instance of the green category, hence the low recognition accuracy score. The accuracy is higher when the stimuli are less similar (left panel), or when the representation are more precise (right panel).

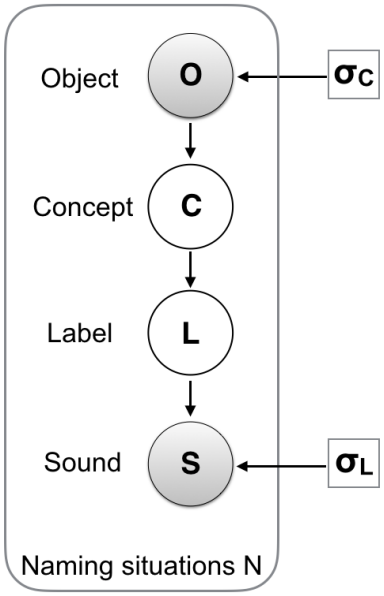

Figure 2. Graphical representation of our model. Circles indicate random variables (shading indicates observed variables). The squares indicate fixed model parameters.

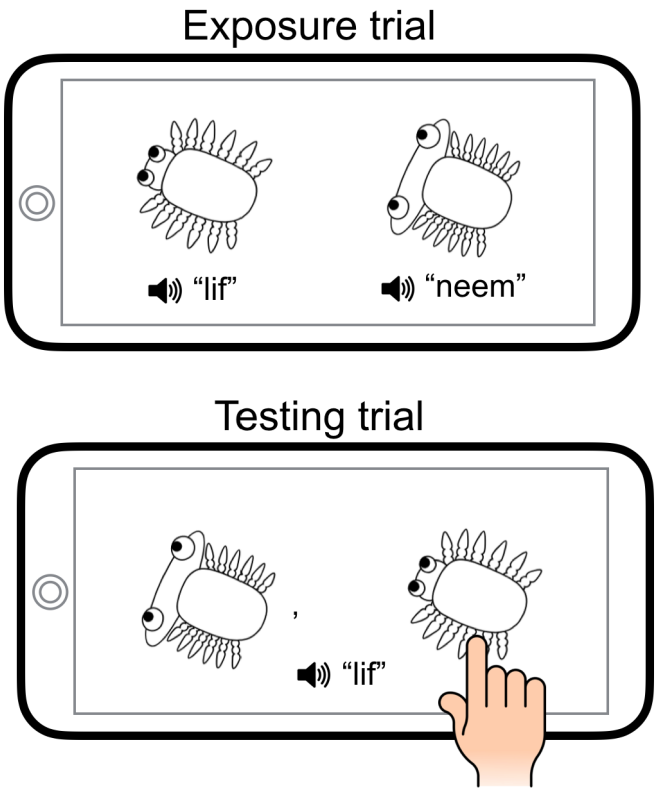

Figure 3. An overview of the task used in this study.

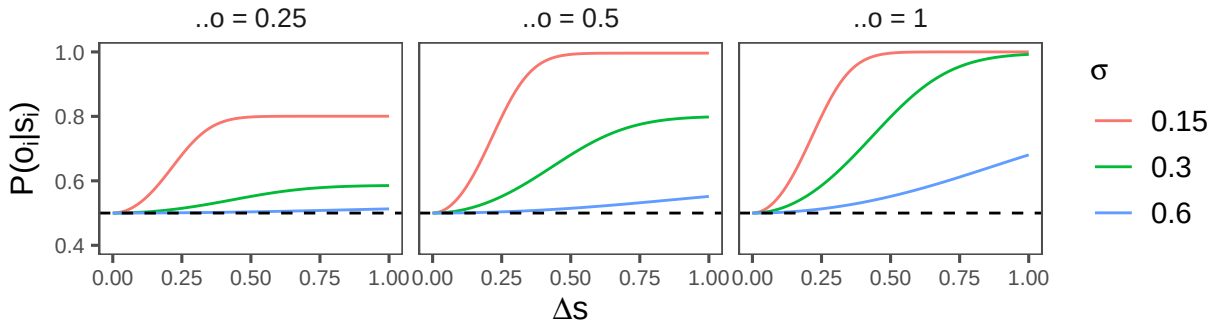

Figure 4. The predicted probability of accurate responses in the testing phase as a function of stimuli distinctiveness Δs and Δo and representation precision σ (For simplicity, we use model 2, which assumes that σ=σ_C=σ_L). Dashed line represents chance.

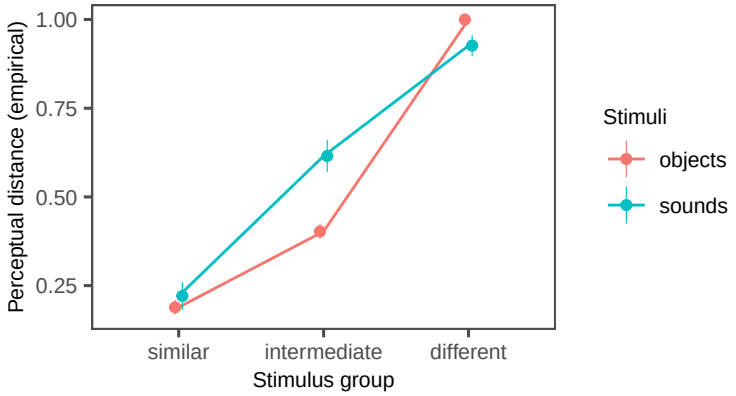

Figure 5. Distances for sound and object pairs (Δs and Δo, respectively) from an adult norming study. Data represent Likert values normalized to [0,1] interval. Error bars represent 95% confidence intervals.

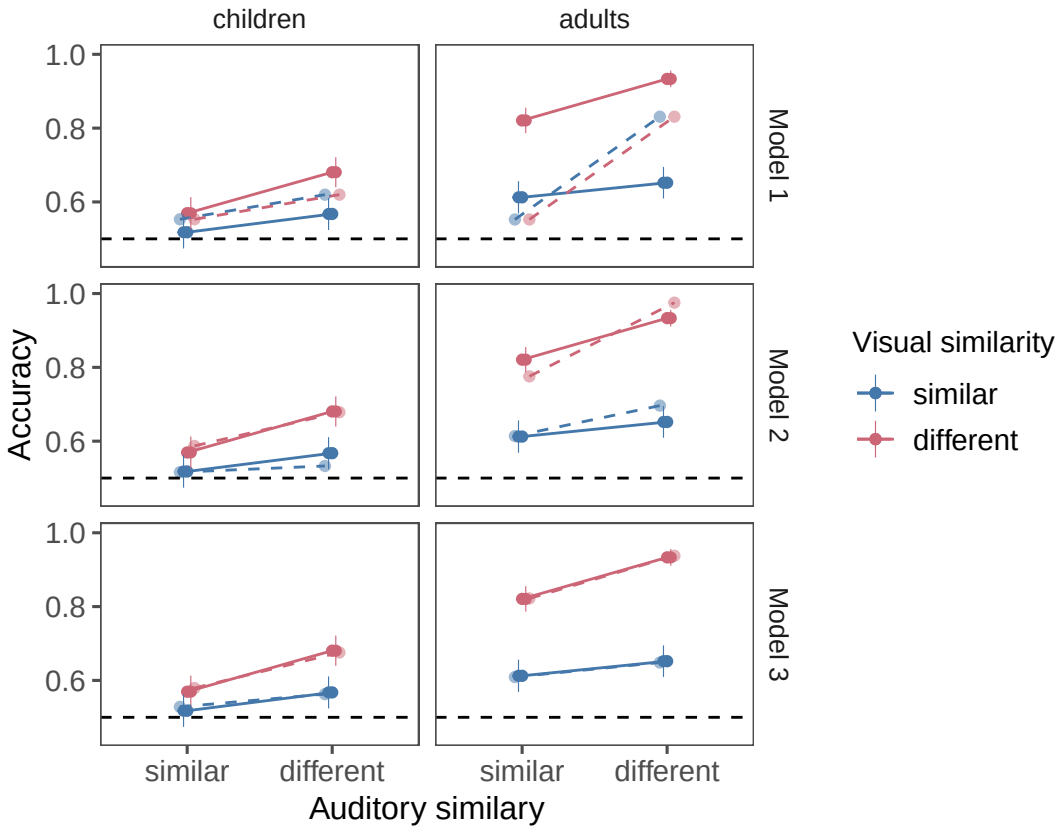

Figure 6. Accuracy of word recognition as a function of the sound distance, the object distance, and the age group (preschool children vs. adults). We show both the models' predictions (dashed lines) and the experimental results (solid lines, same across the three panels). Error bars represent 95% confidence intervals.