Continuous developmental change can explain discontinuities in word learning

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

[1] Department of Psychology, Stanford University

Author Note

Abdellah Fourtassi

Department of Psychology

Stanford University

50 Serra Mall

Jordan Hall, Building 420

Stanford, CA 94301

Correspondence concerning this article should be addressed to Abdellah Fourtassi, Postal address. E-mail: afourtas@stanford.edu

Abstract

"Cognitive development is often characterized in term 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 simpler testing methods) that children do encode such contrasts, thus favoring a continuous pattern of development. Here we propose a probabilistic model describing how development may proceed in a continuous fashion across the lifespan. The model accounts for previously documented facts and provides new predictions. We collected data from preschool children and adults, and we showed that the model can explain various patterns of learning both within the same age and across development. The findings suggest that major aspects 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 can explain discontinuities in word learning

Introduction

Cognitive development is sometimes characterized in terms of a succession of discontinuous stages (Piaget, 1954). Although intuitively appealing, stage theories 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. This appearance can be due, for example, to the use of a cognitively-demanding task which may mask learning, or to the use of statistical thresholding (in particular, p-value < 0.05) 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.

We use a case study from word learning literature. Stager and 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., Word A with Object A and Word B with Object B). Second, they are tested using two types of trials. The control "same" trial consists of a correct pairing (e.g., Word A with Object A) and the "switch" trial consists of a wrong pairing (e.g., Word A with Object B). If babies have correctly learned the association during the familiarization, 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 has initially been 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 (Werker, Fennell, Corcoran, & Stager, 2002), typically do.

The initial discontinuous interpretation has been challenged by subsequent work. For instance, Yoshida, Fennell, Swingley, and Werker (2009) investigated whether failure in the switch task reflects a lack of sound encoding during *familiarization*, or whether it is only due to the demands of the *testing* method which does not allow learning below a certain threshold to be detected. They used the same familiarization procedure as Stager and 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 simpler and finer-grained task, the researcher 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 more challenging (Pajak, Creel, & Levy, 2016; White, Yee, Blumstein, & Morgan, 2013).

Some researchers (Pajak et al., 2016; e.g., Swingley, 2007; Yoshida et al., 2009) proposed that the phonological form 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 the phonological form of words in a probabilistic fashion. According to this view, 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 representations.

In a probabilistic account, a word can be represented as a probability distribution over sound instances organized in a similarity space. The probability is highest at the most typical sound instance. It decreases as the instance becomes less typical. The precision of the representation can be characterized by how much 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.

This framework can explain several findings. In Stager and Werker's original experiment, children are supposed to associate one label "bih" with object 1 and a second label "dih" with object 2. Though infants could learn that the label "bih" is a better match to object 1 than "dih", they could still judge the sound "dih" as a plausible instance of the label "bih", thanks to the relatively large variance of the early encoding, and this confusion leads to "failure" in the recognition task (Figure 1, top). Though learning is small and is easily disrupted by the Switch task, it can still be detected when less demanding methods are used (Yoshida et al., 2009).

The learning accuracy increases (and is detected even by demanding tests such as the Switch) for more distinct word-forms (e.g., "lif" vs. "neem") where the perceptual distance is large relative to the variance (Figure 1, left). 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).

Finally, development can be understood as an increase in the precision (i.e., a decrease in the variance) of the probabilistic representations (Figure 1, right). Such an increase in precision renders minimally different sounds less confusing. Importantly, a more precise representation still has a non-zero variance — Learning difficulties can still be induced with challenging stimuli or in cognitively demanding situations as was demonstrate in adults studies (Pajak et al., 2016; White et al., 2013).

This study

The probabilistic account has been put forward to explain patterns of learning and development at the qualitative level. However, it is crucial to have a precise computational instantiation of this account which would allow us to *quantify* its explanatory power. We could find one previous work that attempted to provide such a computational instantiation (Hofer and Levy, 2017). However, this previous work was designed with the goal of 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 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 (i.e., the effect of the referents' similarity on word learning). Crucially, we explored the extent to which development can be understood 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 learning task which resembles the one used in

Stager and Werker (1997) and Yoshida et al. (2009). 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 the lights of existing accounts of word development.

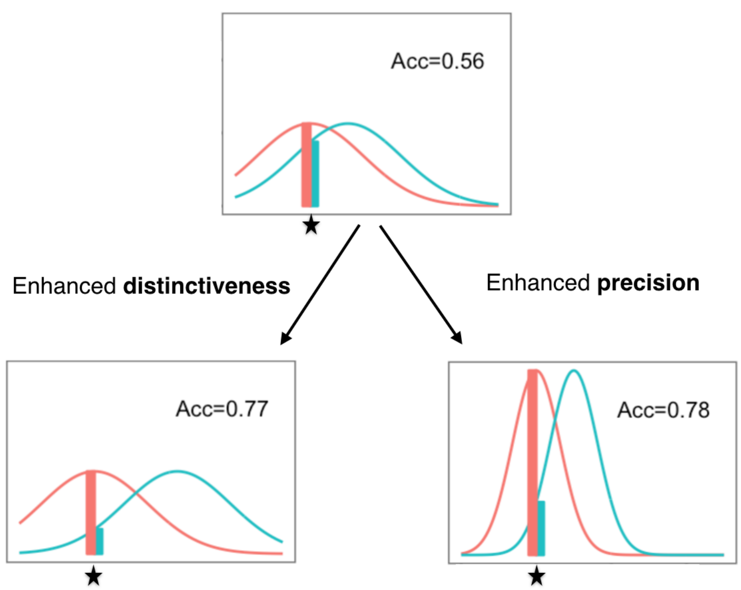

Figure 1. An illustration of the probabilistic/continuous account using simulated data. A word is represented with a distribution over the perceptual space (indicated in red or blue). When the uncertainty of the representation is large relative to the distance between the stimuli (top panel), an instance 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 increases when the stimuli are less similar (left panel), or when the representation are more precise (right panel).

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 and Frank (2013) to model concept learning, and by Hofer and 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 assume that both labels and concepts are observed with a certain amount of perceptual noise, which we assume, for simplicity, is captured by a normal distribution:

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

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

Finally, we assume there to be one-to-one mappings between concepts and labels and that observers have successfully learned these mappings during the exposure phase:

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

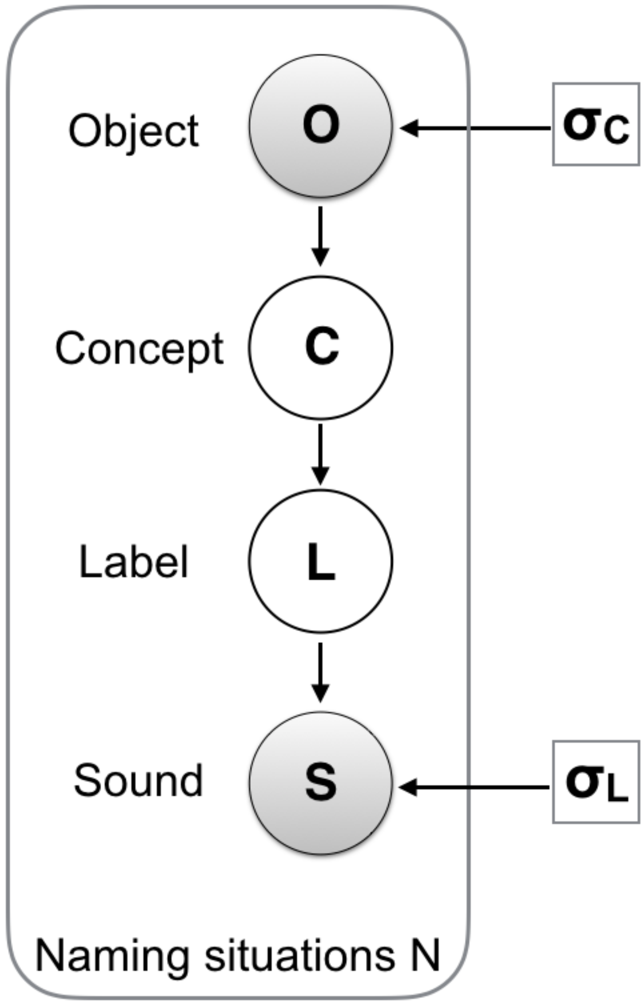

Figure 2. Graphical representation of our model. Circles indicate random variables (shading indicates observed variables). The squares indicate fixed model parameters.

Inference

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:

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

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

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

which can be transformed using Bayes rule into:

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

Finally, assuming that the concepts' prior probability is uniformly distributed[1], we obtain the following expression, where all conditional dependencies are now well defined:

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

Task and model predictions

We use the model to predict word learning in a task similar to the one introduced by Stager and 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 the association between pairs of nonsense words (e.g., "lif"/"neem") and pairs of objects. 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 (1), we derive three exact analytical solutions instantiating different learning assumptions. The first solution is derived by assuming that the labels are

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

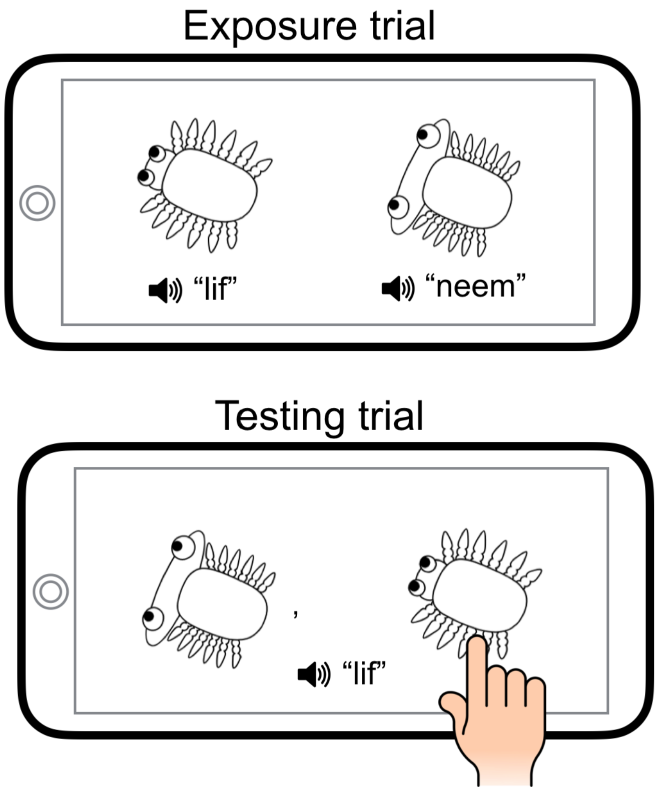

Figure 3. An overview of the task used in this study.

recovered from sounds with a certain level of uncertainty σ_L, 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. One important implication of this assumption is that only the similarity of word sounds modulates success in word learning, not the similarity of the referents (as long as these referents are differentiated perceptually). This assumption yields the following probability function:

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

The second solution is derived by making the more general assumption that both the labels and the concepts are recovered with ambiguity from the sounds and objects. We first 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 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 minimal free parameters.

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

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

$$P(o_T|s) = \frac{1 + e^{-(\frac{\Delta s^2}{2\sigma_L^2} + \frac{\Delta o^2}{2\sigma_C^2})}}{1 + e^{-(\frac{\Delta s^2}{2\sigma_L^2} + \frac{\Delta o^2}{2\sigma_C^2})} + e^{-\frac{\Delta s^2}{2\sigma_L^2}} + e^{-\frac{\Delta o^2}{2\sigma_C^2}}} \tag{3}$$

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

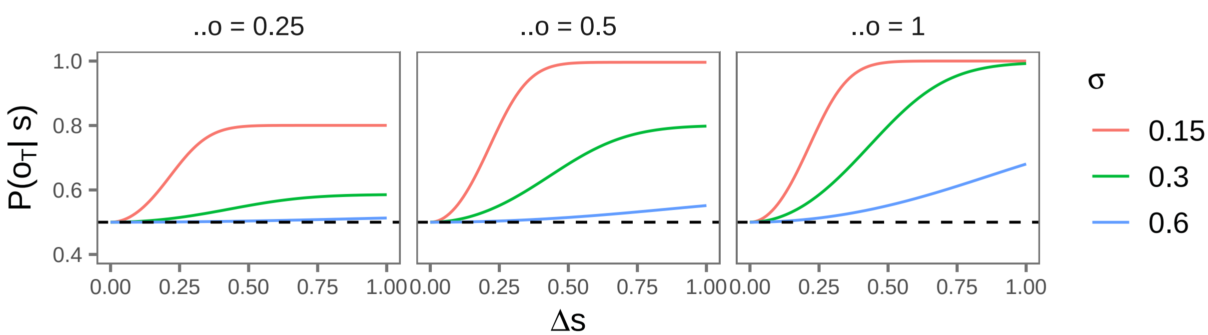

Figure 4. The predicted probability of accurate responses in the testing phase as a function of stimuli distinctiveness Δs and Δo and representation precision σ (For clarity, we assume here that σ=σ_C=σ_L). Dashed line represents chance.

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 fact may explain development, i.e., 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. This is a new prediction that our model makes. Previous work studied the effect of several bottom-up and top-down properties in disambiguating similar sounding words (e.g., Fennell & Waxman, 2010; Rost & McMurray, 2009; Thiessen, 2007), but to our knowledge, no previous study in the literature tested the effect of the visual distance between the semantic referents.

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 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.[2]

[2]https://osf.io/942gv/

Methods

Participants. We report data from $N = 63$ children ages 4-5 years from the Bing Nursery School on Stanford University's campus. An additional $N = 39$ children participated but were removed from analyses because they were not above chance on the catch trials due to the challenging nature of our procedure (see below). We also report data from $N = 74$ adult participants tested on Amazon Mechanical Turk. An additional $N = 26$ were tested but removed from analyses because they had low scores on the catch trials or because they were 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: "ada"/"ad^{h}a" (in hindi) and "aʕa"/"aħa" (in arabic).

As for the objects, we used the Dynamic Stimuli javascript library[3] which allowed us to generate objects in four different categories: "tree," "bird," "bug," and "fish." These categories are supposed to be naturally occurring kinds that might be seen 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 $N = 20$ adults 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. These data will be

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

used in all models as a proxy for the perceptual distance between the sound pairs (Δs) and the object pairs (Δo).

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. *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 through using the appropriate distance (Figure 5).

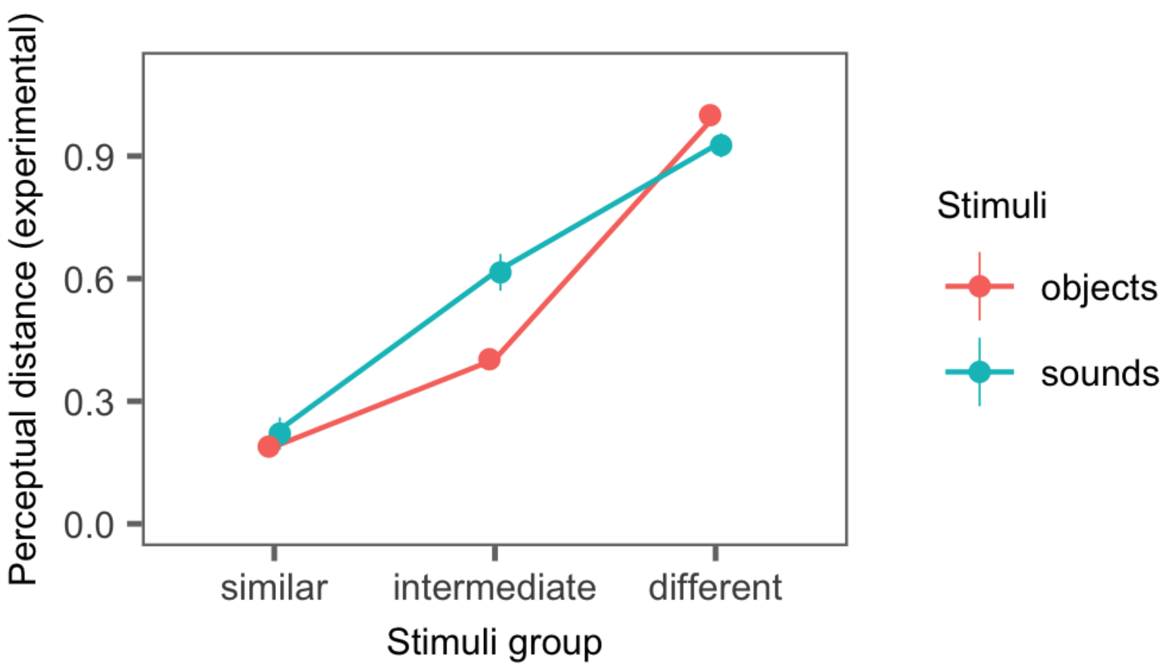

Figure 5. Distances for both sound and object pairs from an adult norming study. Data represent Likert values normalized to [0,1] interval. Error bars represent 95% confidence intervals.

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 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 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). 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 $\beta = 0.52$ ($p < 0.001$), replicating previous findings that sound distance modulates success in word learning (e.g., Stager & Werker, 1997). For object distance, we found $\beta = 0.83$ ($p < 0.001$), and this finding confirms the new prediction of our model according to which object distance also modulates success in word learning. Finally, for the age group, we obtained $\beta = 0.76$ ($p < 0.001$), showing that overall performance improves with age.

Table 1

Characteristics and performance of the models used in this study.

				Children		Adults	
Model	Structure	Param.	R^2	Sig_L	Sig_C	Sig_L	Sig_C
model 1	Sig_L only	1	0.27	1.00	–	0.37	–
model 2	$Sig_L = Sig_C$	1	0.95	0.60	0.6	0.15	0.15
model 3	Sig_L != Sig_C	2	1.00	0.83	0.31	0.12	0.17

We next fit the three models obtained through expressions (1), (2), and (3) to the participants' responses in each age group. The predictions of the models are shown 6 (dashed lines) and 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, suggesting that the explanatory power of the general Model 3 is largely due to its structure, rather than its degrees of freedom.

As predicted, the uncertainty parameters were larger for children than they were for adults (Table 1), showing that the probabilistic representations become more refined (that is, σ becomes smaller) across development. Further, the parameter estimates of Model 3 show that this developmental effect is more important for the label-specific uncertainty that it is for the concept-specific uncertainty.

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 and Werker (1997) showing a discrepancy between children's speech perception abilities and their word learning skills. While this fact might seem like a specific stage in early development, our model demonstrated, instead, that it can be more simply understood in terms of continuous developmental change.

The main assumption of the model was that both word form and referent are encoded in a probabilistic fashion. The model provided a quantitative instantiation of the continuous development hypothesis (Pajak et al., 2016; e.g., Swingley, 2007; Yoshida et al., 2009)). More precisely, we showed that developmental changes in word-object mappings can be characterized as a continuous refinement in the precision of the probabilistic representations.

We find in the literature two broad accounts of development in the Switch task: 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, i.e., contrasting word meaning (Apfelbaum & McMurray, 2011; Dietrich, Swingley, & Werker, 2007; Rost & McMurray, 2009, 2010; 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. These two accounts are complementary and both seem to play a role (see Tsui, Byers-Heinlein, & Fennell, 2019 for a review).

Our model is compatible with both of these accounts. In our work, the probability distributions do not distinguish between the direct and indirect sources of uncertainty — both are included. Indeed, part of the measured uncertainty reflects the learner's degrees of confidence in the phonetic/phonological boundaries (i.e., the direct account) and another part reflects a possible drop in perceptual acuity due to high cognitive load (i.e., the indirect account). Note, however, that the model (at least in its current format) is incapable of answering questions about the development of each of these sources of uncertainty separately or about their relative contribution to the global uncertainty.

Werker and Curtin (2005) proposed to explain development in the Switch task using their theory called Processing Rich Information from Multidimensional Interactive Representations (or PRIMIR) which purports to explain various phenomena in early speech perception and word learning within a unified framework. PRIMIR posits that children initially try to attend to various features of the speech signal, regardless of whether or not these features are relevant to the task at hand. For example, when learning the meaning of similar sounds, infants are unsure what detail is most important to identify words (i.e., the phonemes), and will instead activate several aspects of the information simultaneously (including, for example, the gender of the speaker). The lack of attentional selection leads to confusion and then to failure in the task.

According to PRIMIR, learning similar-sounding words becomes more robust over time as children develop abstract phonemic categories. The latter act as filters, allowing children to attend selectively to the important information. This account is also compatible with our model (as it resembles the direct account we discussed above). Developing a better attentional strategy (thanks to phoneme) can translate into a reduction in the uncertainty about whether a sound contrast signals a change in meaning.

While most research focused on the sound representation in analyzing the Switch task, 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.

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. More precisely, a comparison of the variance estimates across age groups shows 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 does. 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.

Our finding that word learning is mediated by the visual similaérity of the semantic objects has implications for theories of lexical development. It suggests that, all things being equal, children may prioritize the acquisition of 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. This suggestion is, indeed, supported by recent work investigating vocabulary development (Engelthaler & Hills, 2017; Fourtassi, Bian, & Frank, 2018; Sizemore, Karuza, Giusti, & Bassett, 2018)

There are a few 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 between-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 differences compared to adults, suggesting that development does not stop at 17 months, but continues throughout childhood.

One limitation of the model is that it only accounts for bottom-up, similarity-based effects. It does not account for how high-level factors such as social and communicative cues can influence learning. For example, Fennell and Waxman (2010) highlighted the fact that the Switch task introduces 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.

All data and code for these analyses are available at
https://github.com/afourtassi/networks

Acknowledgements

This work was supported by a post-doctoral grant from the Fyssen Foundation, NSF #1528526, and NSF #1659585.

Disclosure statement

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

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.

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., & 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 *CogSci*.

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*.

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

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*.

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. (2007). Lexical exposure and word-form encoding in 1.5-year-olds.

Developmental Psychology, *43*(2).

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., & 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., 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., 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).

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*.

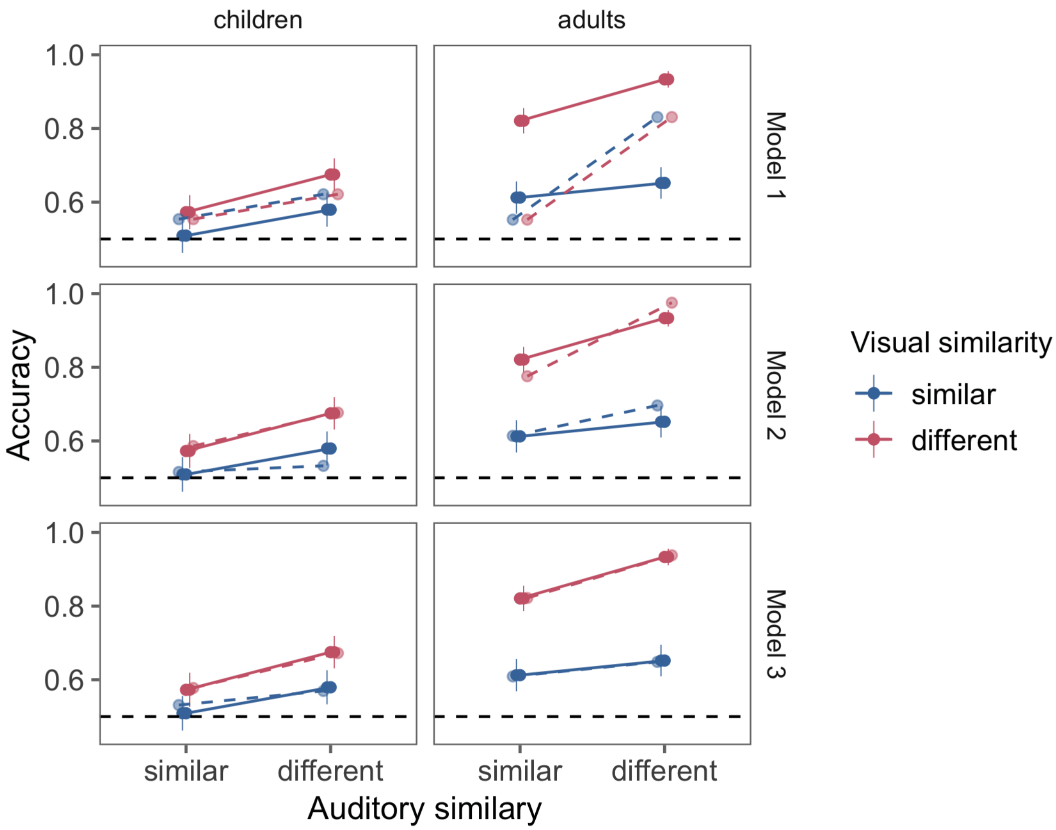

Figure 6. Accuracy of word recognition as a function of the sound distance, the object distance, and the age group (preschool children vs. adults). We show both the models' predictions (dashed lines) and the experimental results (solid lines). The single-variance model uses one joint fitting parameter for both sound and meaning variances. The double-variance model uses two separate fittings parameters for the sound and the meaning variances. Error bars represent 95% confidence intervals.