

Continuous developmental change can explain discontinuities in word learning

Abdellah Fourtassi¹, Sophie Regan¹, & Michael C. Frank¹

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, $p\text{-value} < 0.05$) 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. Indeed, though infants around 14-month old can distinguish similar sound pairs such as “dih” and “bih”, they appear to fail in mapping this pair to two different objects.

By around 17 m.o, children succeed in the same task (Werker, Fennell, Corcoran, & Stager, 2002). How does development proceed? The answer to this question depends on the nature of the children’s word encoding in the task. Early accounts assumed that children encode words in a rather binary way: they either fail or succeed in attending to the relevant

phonetic details (simultaneously with the meanings). In the original paper, Stager and Werker (1997) noted that “Infants of 14 months fail to detect the same phonetic detail that they can easily detect in a simple syllable discrimination task”. They argued that the high cognitive demands of the task make children experience a limitation in the amount of phonetic details they can access. In a similar vein, Pater et al. (2004) argued that “when a contrast is first acquired, it is not stable and can be lost under processing demands”. This binary account suggested a discontinuous/stage-like pattern of development whereby younger children fail to encode the contrastive phonetic detail, whereas older children succeed.

Subsequent findings have investigated the implications of this hypothesis in both infants and adults. First, if the demands of the task totally impede the encoding of subtle phonetic contrasts, then infants should fail in learning the mappings between the sounds and the objects, regardless of how this learning is probed. However, 14-month-olds succeed in this task under the same *learning* conditions as in Stager and Werker (1997), but when an easier *testing* method is used (Yoshida, Fennell, Swingley, & Werker, 2009). Second, if the mismatch between sound discrimination and word learning is only related to limited cognitive resources in infancy, then adults — who have greater attentional and working memory capacity — should always succeed in mapping pairs of sounds they can discriminate to different objects. Nonetheless, even adults show patterns learning that mirror those shown by 14-month-olds (Pajak, Creel, & Levy, 2016; White, Yee, Blumstein, & Morgan, 2013).

This new set of evidence points towards another scenario, where the representations are encoded in a probabilistic (rather than binary) way, and where development is continuous, rather than stage-like (see also Swingley, 2007). On this account, correct representations are learned early in development, but these representations are encoded with higher uncertainty in younger children, leading to *apparent* failure in relatively demanding tasks. Development is a continuous process whereby the initial noisy representations become more precise. Crucially, in a probabilistic account, more precise representation do not mean

they perfect, thus accounting for the fact that even adults show low accuracy learning when the sounds are subtle, e.g., non-native (Pajak et al., 2016).

We provide an intuitive illustration of how such an account explains patterns of learning and development in Figure 1. We observe low accuracy in word learning when the perceptual distance between the labels is small relative to the uncertainty with which these labels are encoded. For example, in Stager and Werker’s original experiment, children are supposed to associate label 1 (“bih”) and label 2 (“dih”) with object 1 and object 2, respectively. 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 uncertainty of the encoding, and this confusion leads to “failure” in the recognition task. According to this account, accuracy in word learning improves if we increase either the perceptual distinctiveness of the stimuli (e.g., through using different-sounding labels), or the precision of the encoding itself (e.g., across development).

Building on this intuition, the current work proposes a probabilistic model, which we use to both account for previous experimental findings, and to make new predictions that have not been tested before. Using new data collected from both preschool children and adults, we show that the model can explain various patterns of learning both within the same age and across 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

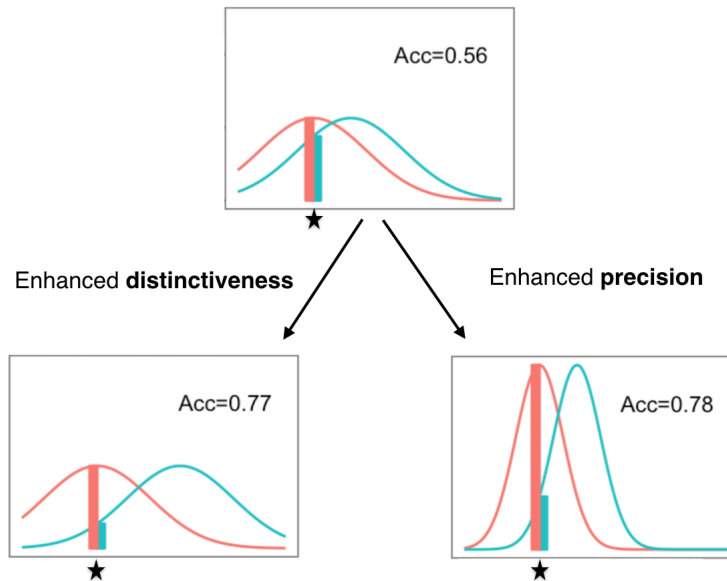


Figure 1. An illustration of the probabilistic/continuous account using simulated data. A word is represented with a distribution over the perceptual space (indicated in red or blue). When the uncertainty of the representation is large relative to the distance between the stimuli (top panel), an instance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

104 concept C in the speaker’s mind. The concept prompted the corresponding label L . Finally,
 105 the label was physically instantiated by the sound s .

106 A similar probabilistic structure was used by Lewis and Frank (2013) to model concept
 107 learning, and by Hofer and Levy (2017) to model spoken word learning. However, the first
 108 study assumed that the sounds are heard unambiguously, and the second assumed the
 109 concepts are observed unambiguously. In our model, we assume that both labels and
 110 concepts are observed with a certain amount of perceptual noise, which we assume, for
 111 simplicity, is captured by a normal distribution:

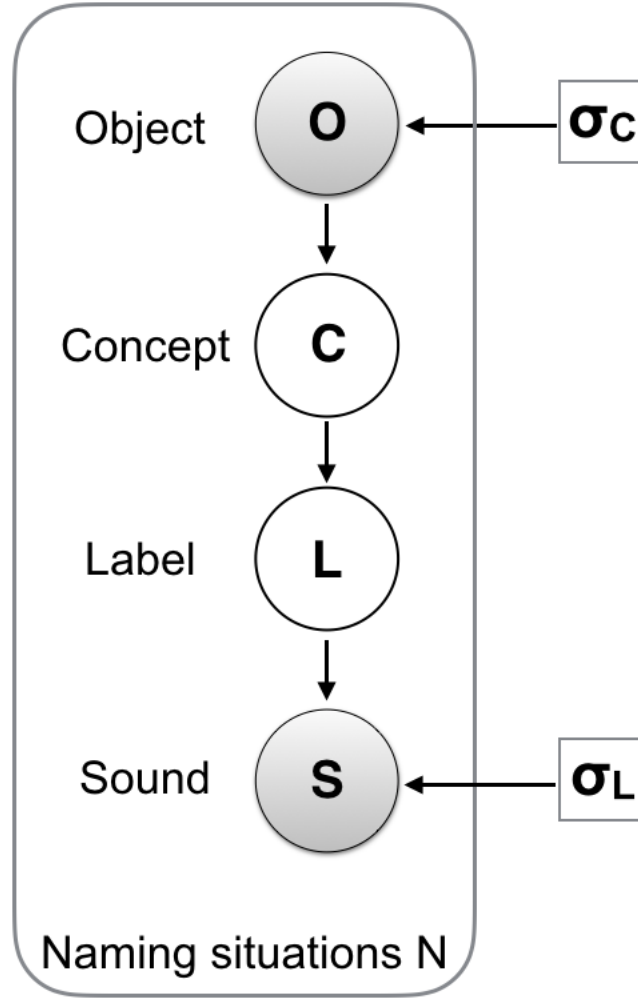


Figure 2. Graphical representation of our model. Circles indicate random variables (shading indicates observed variables). The squares indicate fixed model parameters.

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

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

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

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

114 Inference

115 The learner hears a sound s and has to decide which object o provides an optimal
 116 match to this sound. To this end, they must compute the probability $P(o|s)$ for all possible
 117 objects. This probability can be computed by summing over all possible concepts and labels:

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

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

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

120 which can be transformed using Bayes rule into:

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

121 Finally, assuming that the concepts' prior probability is uniformly distributed¹, we
 122 obtain the following expression, where all conditional dependencies are now well defined:

$$P(o|s) = \frac{\sum_{C,L} P(s|L)P(o|C)P(L|C)}{\sum_o \sum_{C,L} P(s|L)P(o|C)P(L|C)} \quad (1)$$

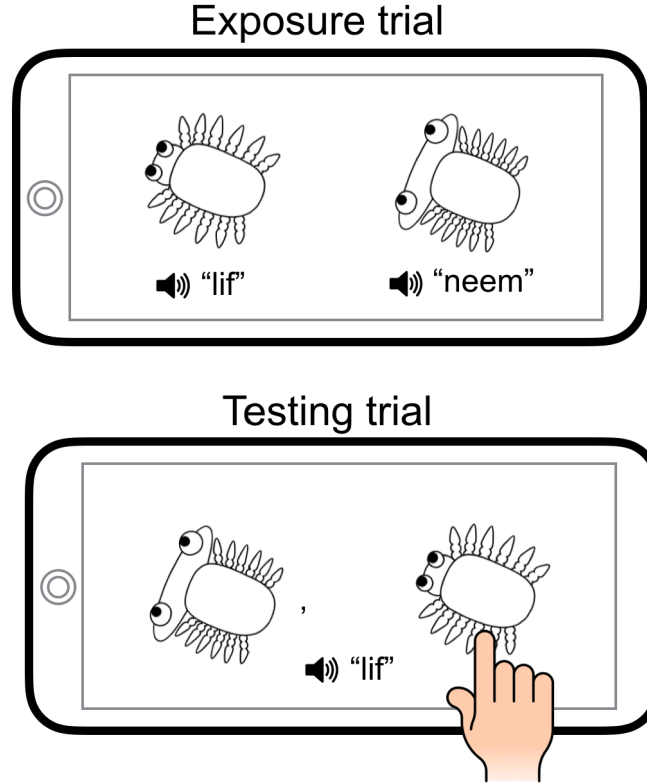


Figure 3. An overview of the task used in this study.

Task and model predictions

We use the model to predict performance in the word learning task introduced by Stager and Werker (1997), with a two-alternative forced choice as in 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.

We used Equation 1 and the probability distributions defined above to obtain the exact

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

132 analytical expression for the probability of accurate responses $p(o_T|s)$ (target object o_T given
 133 a sound s) in the simple case of two-alternative forced choice in the testing phase of our
 134 experimental task:

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

135 Figure 4 show simulations of the predicted accuracy (Expression 2) as a function of the
 136 distinctiveness parameters (Δs and Δo) and the precision parameters, i.e., the variances of
 137 the distributions $p(s|L)$ and $p(o|C)$. To understand the qualitative behavior of the model, we
 138 assumed for simplicity that the precision parameter has similar values in both distributions,
 139 i.e., $\sigma = \sigma_C \approx \sigma_L$ (but we will allow those parameters to vary independently in the rest of
 140 the paper).

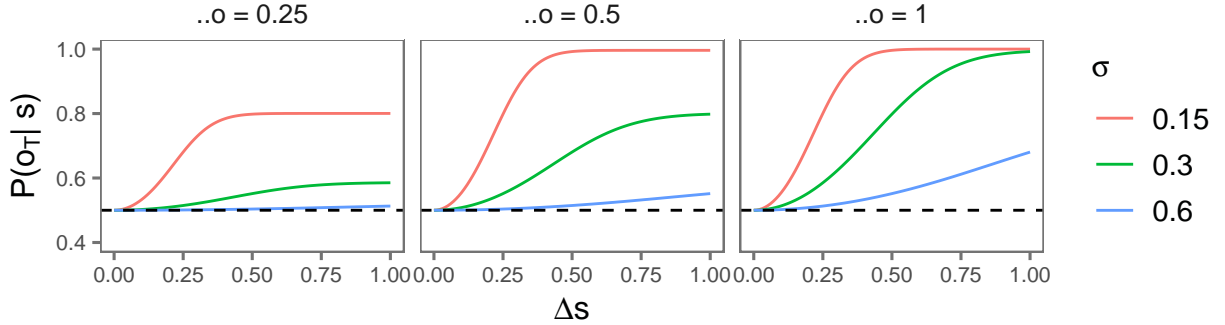


Figure 4. The predicted probability of accurate responses in the testing phase as a function of stimuli distinctiveness Δs and Δo and representation precision σ (For clarity, we assume here that $\sigma = \sigma_C = \sigma_L$). Dashed line represents chance.

141 The simulations explain some previously documented facts, and make new predictions:

- 142 1) For fixed values of Δo and σ , the probability of accurate responses increases as a
 143 function of Δs . This pattern accounts for the fact that similar sounds are generally
 144 more challenging to learn than different sounds for both children (Stager & Werker,
 145 1997) and adults (Pajak et al., 2016).

- 2) For fixed values of Δs and Δo , accuracy increases when the representational uncertainty (characterized with 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.

Methods

Participants. We planned to recruit a sample of $N=60$ children ages 4-5 years from the Bing Nursery School on Stanford University's campus. Here we report data from $n=55$ children. An additional $n=35$ children participated but were removed from analyses because they were not above chance on the catch trials due to the challenging nature of our

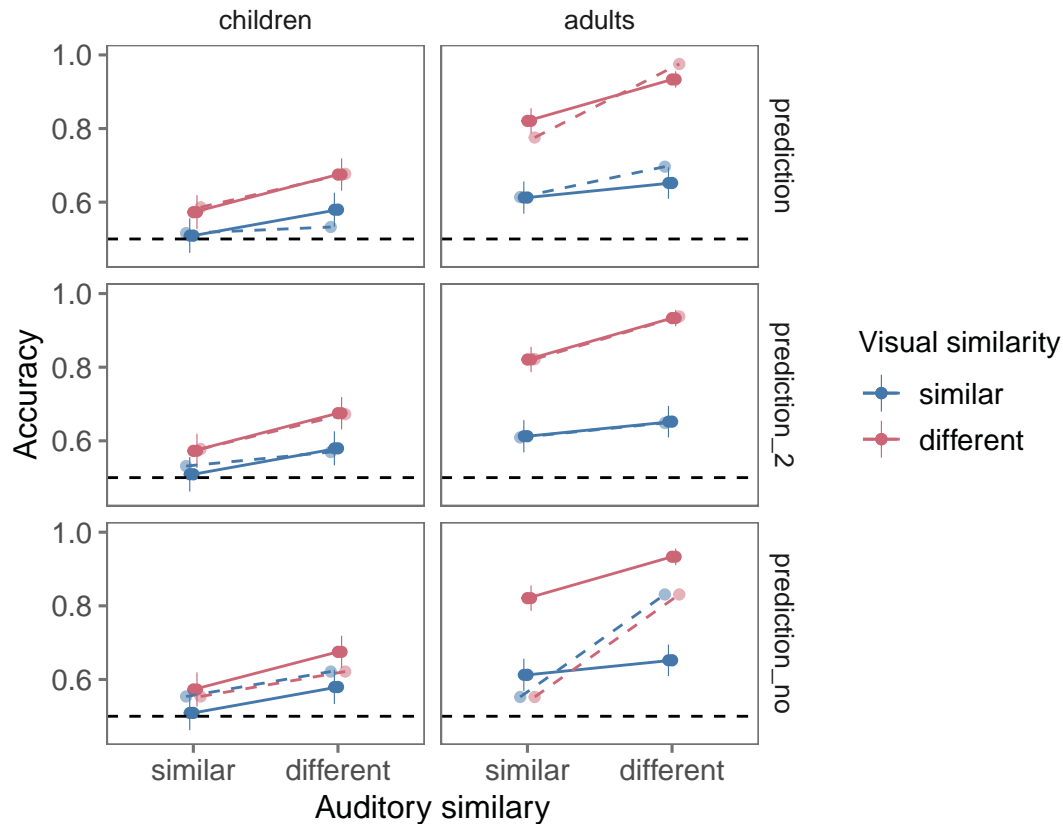
procedure (see below). We also planned to recruit a sample of $N=60$ adults on Amazon Mechanical Turk. Data from $n=26$ participants were excluded due to low scores on the catch trials ($n=26$) or because they were familiar with the non-English sound stimuli we used in the adult experiment ($n=0$), yielding a final sample of $n=74$.

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 similarity to English speakers: 1) “different”: “lif”/“neem” and “zem”/“doof”, 2) “intermediate”: “aka”/“ama” and “ada”/“aba”, and 3) “similar” non-English minimal pairs: “ada”/“ad^ha” (in hindi) and “aʃa”/“aḥa” (in arabic).

As for the objects, we used the Dynamic Stimuli javascript library² 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 a separate survey, $N = 20$ participants recruited on Amazon Mechanical Turk evaluated the similarity of each sound and object pair on a 7-point scale. We scaled responses within the range $[0,1]$. Data are shown in Figure 5, for each stimulus group. These data will be used in the models as the perceptual distance of sound pairs (Δs) and object pairs (Δo).

²<https://github.com/erindb/stimuli>



###Design

Each age group saw only two of the three levels of similarity described in the previous sub-section: “different” vs. “intermediate” for 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 perceptual distance used in each age group (Figure 5).

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 4 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.

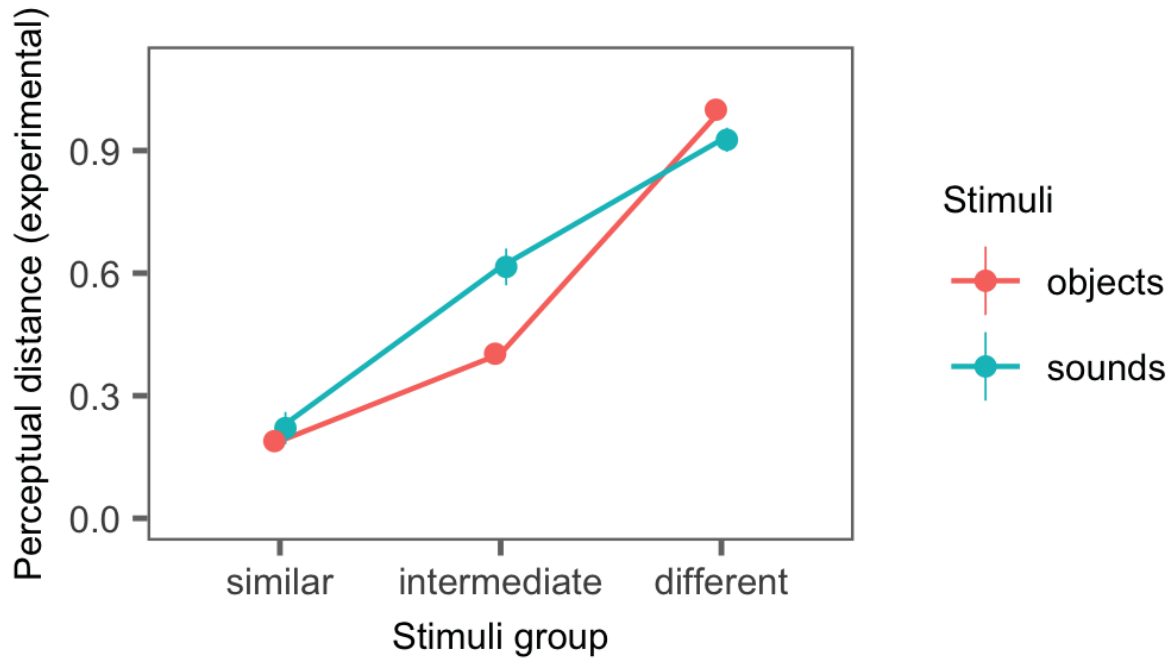


Figure 5. Distances for both sound and object pairs from an adult norming study. Data represent Likert values normalized to $[0,1]$ interval. Error bars represent 95% confidence intervals.

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. For object distance, we found $\beta = 0.83$ ($p < 0.001$), and this finding confirms the new prediction of our model. Finally, for the age group, we obtained $\beta = 0.76$ ($p < 0.001$), showing that performance improves with age. In addition, we found two-way interactions between sound distance and age ($\beta = 0.15$, $p = 0.17$) and between object distance and age ($\beta = 0.45$, $p < 0.001$).

We next fit our model (using Equation 2) to the participants’ responses in each age group. The values of Δs and Δo were set based on data from the similarity judgment task (Figure 5). Thus, the model has two degrees of freedom for each group, i.e., σ_C and σ_L . Figure 6 (dashed lines) shows the predictions. The model captures the qualitative patterns

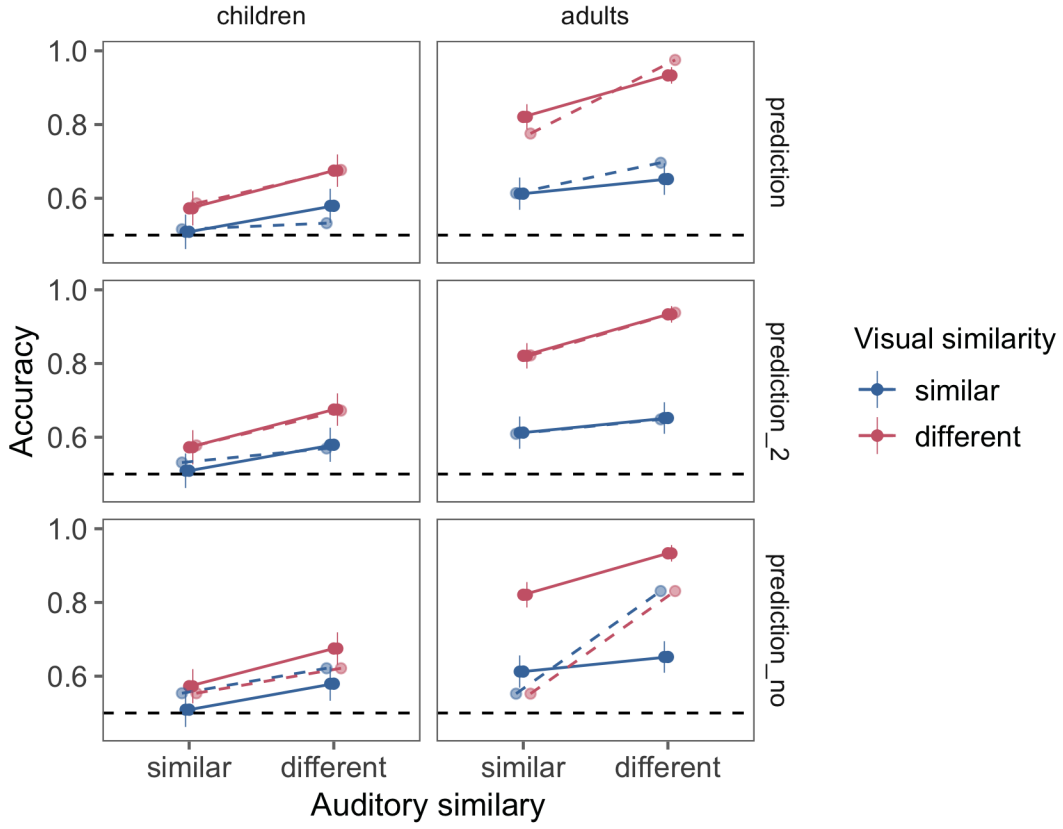


Figure 6. Accuracy of novel word recognition as a function of the sound distance, the object distance, and the age group (preschool children vs. adults). We show both experimental results (solid lines) and model predictions (dashed lines). Error bars represent 95% confidence intervals.

in both age groups: starting from a low accuracy recognition when both the sound and object distances are small, the model correctly predicts an increase in accuracy when either the sound distance or the object distance increases. Further, accuracy is correctly predicted to be maximal when both the sound and object distances are high.

The values of the parameters were as follows. Children had a label-specific uncertainty of $\sigma_S = 0.83$ [0.64, 1.02]³, and a concept-specific uncertainty of $\sigma_C = 0.31$ [0.11, 0.51]. Adults had a label-specific uncertainty of $\sigma_S = 0.12$ [0.12, 0.13], and a concept-specific uncertainty

³All uncertainty intervals in this paper represent 95% Confidence Intervals.

of $\sigma_C = 0.17$ [0.16, 0.18]. As predicted, the uncertainty parameters were larger for children than they were for adults, showing that the probabilistic representations becomes more refined (that is, σ becomes smaller) across development. The developmental effect was more important for the label-specific uncertainty.

The models explained the majority of the variance in the participants' mean responses ($R^2 = 1$, for the combined adult and children's data). To investigate whether the model's predictive power was due to overfitting, we fit a simplified version with only one degree of freedom (i.e., a single variance common to both sounds and objects). This single-variance model explained as much variance in the mean responses ($R^2 = 0.95$). It also captured the main qualitative patterns (graph not shown), suggesting that the explanatory power of the model is largely due to its structure, rather than its degrees of freedom.

An unexpected outcome was that adult participants deviated slightly from the model's numbers: While the model predicted accuracy to be more sensitive to object distance when sound distance is higher (and vice-versa), adult participants showed the opposite pattern. This deviation is an interesting starting point for future work as it may suggest that participants are more likely to pay attention to and integrate additional sources of information when ambiguity is higher.

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 much of the previous investigation of this finding has been interested in the source of this discrepancy, here we have explored how it could arise from continuous developmental change in perceptual uncertainty.

Building on some previous discussions (e.g., Swingley, 2007; Yoshida et al., 2009), we proposed a model where perceptual stimuli are encoded probabilistically. We tested the model’s predictions against data collected from preschool children and adults and we showed that developmental changes in word-object mappings can indeed be characterized as a continuous refinement (i.e., uncertainty reduction) in qualitatively similar representations across the life span.

The model made a new prediction to which we tested experimentally: Learning similar words is not only modulated by the similarity of their phonological forms, but also by the visual similarity of their semantic referents. More generally, since visual similarity is an early organizing feature in the semantic domain (e.g., Wojcik & Saffran, 2013), our finding suggests that children may prioritize the acquisition of words that are quite distant in the semantic space. This suggestion is supported by recent findings based on the investigation of early vocabulary growth (Engelthaler & Hills, 2017; Sizemore, Karuza, Giusti, & Bassett, 2018).

One limitation of this work 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 (14-18 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.

In sum, 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 development, 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

- 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.
- 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

311 distinctiveness predicts age of acquisition. *Cognitive Science*, 41.

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

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

316 Lewis, M., & Frank, M. (2013). An integrated model of concept learning and word-concept
317 mapping. In *Proceedings of the annual meeting of the cognitive science society* (Vol.
318 35).

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

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

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

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

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

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

331 Thiessen, E. (2007). The effect of distributional information on children's use of phonemic

332 contrasts. *Journal of Memory and Language*, 56.

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

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

337 Wojcik, E., & Saffran, J. (2013). The ontogeny of lexical networks: Toddlers encode the
338 relationships among referents when learning novel words. *Psychological Science*,
339 24(10).

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