



Cognitive Science 41 (2017, Suppl. 1) 141–161
Copyright © 2015 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12322

Developmental Changes in Cross-Situational Word Learning: The Inverse Effect of Initial Accuracy

Stanka A. Fitneva,^a Morten H. Christiansen^b

^a*Department of Psychology, Queen's University*

^b*Department of Psychology, Cornell University*

Received 12 February 2015; received in revised form 28 July 2015; accepted 4 August 2015

Abstract

Intuitively, the accuracy of initial word-referent mappings should be positively correlated with the outcome of learning. Yet recent evidence suggests an inverse effect of initial accuracy in adults, whereby greater accuracy of initial mappings is associated with poorer outcomes in a cross-situational learning task. Here, we examine the impact of initial accuracy on 4-year-olds, 10-year-olds, and adults. For half of the participants most word-referent mappings were initially correct and for the other half most mappings were initially incorrect. Initial accuracy was positively related to learning outcomes in 4-year-olds, had no effect on 10-year-olds' learning, and was inversely related to learning outcomes in adults. Examination of item learning patterns revealed item interdependence for adults and 4-year-olds but not 10-year-olds. These findings point to a qualitative change in language learning processes over development.

Keywords: Language development; Cross-situational learning; Disfluency; Statistical learning; Prediction-based learning

1. Introduction

Existing research suggests that statistical learning mechanisms operate continuously over the lifespan (e.g., Saffran, Aslin, & Newport, 1996; Saffran, Newport, Aslin, Tunick et al., 1997; Smith & Yu, 2008; Yu & Smith, 2007; for reviews, see Frost, Armstrong, Siegelman, & Christiansen, 2015; Krogh, Vlach, & Johnson, 2013; Misyak, Goldstein, & Christiansen, 2012). Children and adults are rarely compared in the same study, however, leading to a number of questions. Beyond continuity in the *existence* of statistical learning skills, what can we say about the stability of their application, functioning, capacity,

Correspondence should be sent to Stanka A. Fitneva, Department of Psychology, Queen's University, Kingston, ON K7L 4V1, Canada. E-mail: fitneva@queensu.ca

and the mechanisms behind them? Experimental data and computational modeling suggest that statistical learning performance is affected by memory, attention, and prior knowledge (Finn & Hudson Kam, 2008; Smith, Suanda, & Yu, 2014; Toro, Sinnett, Soto-Faraco, & Soto-Faraco, 2005). These factors undergo changes of their own over the life span (possibly as a result of statistical learning; Smith et al., 2014), suggesting that both quantitative and qualitative developmental changes in statistical learning are likely. Nevertheless, little research has directly examined potential changes in statistical learning over development other than in closely spaced age groups (Smith & Yu, 2008, 2013; Vlach & Johnson, 2013), which may have obscured the existence of qualitative change.

The goal of this study was to chart performance in cross-situational word learning from preschool to adulthood under closely matched conditions. Cross-situational word learning, extensively investigated in the groundbreaking work of Linda Smith and colleagues (Smith & Yu, 2008; Smith et al., 2014; Yu & Smith, 2007; Yurovsky, Fricker, Yu, & Smith, 2014), refers to finding the meaning of words by tracking the co-occurrences of words and referents across situations. This study aimed to contribute to establishing the similarities and differences in this type of statistical learning across the ages and thus to establishing a framework for research into the development of statistical learning skills. In particular, the study focused on the role of the (in)accuracy of initial word-referent associations in cross-situational learning (Fitneva & Christiansen, 2011).

1.1. The value of accuracy

Vibrant research on how children converge on the correct referent of a word was stimulated by Quine's (1960) framing of the "indeterminacy of reference" problem, viz. individual naming situations are highly ambiguous. As a result, we now know a lot about the conceptual biases that children may have (Markman, 1992) and the linguistic, cognitive, and social cues (e.g., Diesendruck, Carmel, & Markson, 2010; Fitneva, Christiansen, & Monaghan, 2009; Hollich et al., 2000; Regier, 2005; Samuelson & Smith, 1998; Smith, 2000) that separately or together may help children narrow the space of potential referents and identify the correct one. Following Quine's definition of the problem, however, this research focused on single-trial learning (or learning in multiple but identical naming situations, which is methodologically more common). Clearly, errors in this setting are catastrophic to learning.

Recent research questions whether single-trial learning research captures long-term word-referent retention or the situational disambiguation skills of the learner (Bion, Borovsky, & Fernald, 2013; Horst & Samuelson, 2008; McMurray, Horst, & Samuelson, 2012). Focusing on cross-situational—over time—learning, recent research also acknowledges the difficulty posed by referential indeterminacy but questions how important initial errors are given that by encountering a novel word in various situations learners can quickly converge on the target associations (e.g., McMurray et al., 2012; Smith & Yu, 2008; Yu & Smith, 2007). Indeed, both infants and adults show an ability to track and compare information across situations and quickly identify word meanings (Smith & Yu, 2008; Yu & Smith, 2007).

The most prominent models of cross-situational learning advance hypothesis testing and associative learning as the mechanisms underlying it. According to the hypothesis-testing account (Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell, Medina, Hafri, & Gleitman, 2013; see also Frank, Goodman, & Tenenbaum, 2009; Vouloumanos & Werker, 2009), learners form hypotheses about the word-referent associations, which they maintain until proven wrong, at which time they start anew but choosing a referent from a smaller set of alternatives. According to the associationist account (Yu & Smith, 2007), learners track the co-occurrences of words and all possible referents. The frequency of these co-occurrences defines the learning outcome. Although the empirical evidence and theoretical models appear to ask us to move beyond these extreme propositions (Dautriche & Chemla, 2014; Yu & Smith, 2012), it is notable that the probability of making a correct initial mapping appears to be a critical determinant for the outcome of cross-situational learning in both models. There are at least two reasons for this. First, a wrong mapping requires correction on a subsequent trial, which is effortful and disruptive, or it could be encoded in memory and lead to errors in performance. Second, a correct initial mapping may facilitate subsequent learning by supporting the discovery of other target mappings (by narrowing the possibilities for those mappings). Thus, even though erroneous mappings are not necessarily a catastrophic impediment to cross-situational learning, in current models the accuracy of initial word-referent associations underlie its speed and efficiency.

1.2. Good things about errors

Although intuitively it makes sense that accurate initial mappings would set learners on a fast track to vocabulary knowledge, this might not necessarily be the case. Indeed, studies in neural network learning suggest that under certain circumstances initial correct guesses can leave the system stranded in local minima from which it cannot easily escape (Gori & Tesi, 1992; Plagianakos, Magoulas, & Vrahatis, 2001). In addition, although correct initial biases in combination with supporting subsequent evidence could be sufficient for learning (e.g., Xu & Tenenbaum, 2007), learning may be more efficient with errors because errors clearly rule out alternatives. Finally, incorrect initial word-referent mappings may facilitate learning by creating surprise or attracting attention when their inaccuracy becomes apparent. A growing number of findings suggests that experiencing disfluency—or incidental difficulty—in a cognitive task may lead to more elaborate and systematic processing and deeper encoding of information (e.g., Alter, Oppenheimer, Epley, & Eyre, 2007; Mueller & Oppenheimer, 2014; Oppenheimer, 2008; Oppenheimer & Frank, 2007).

1.3. Inverse effect of initial accuracy in word learning

A recent study suggests that incorrect initial pairing of words and referents may facilitate word learning (Fitneva & Christiansen, 2011). In the study, English-speaking adults had to learn the meaning of 10 words within a cross-situational paradigm. Two pictures

and their labels were presented without there being a systematic relation between the location of the referents on the screen and the order in which the words were played. Each word-referent pair was presented four times in the course of the study. Fitneva and Christiansen (2011) used the longest fixation on a picture during the first presentation of the words to estimate the strongest associations made. After assessing the accuracy of these associations with respect to the target mappings, they found that learners with more incorrect initial associations were more accurate at test than learners with more correct initial associations. In other words, initial accuracy was inversely related to learning, perhaps because of greater disfluency experienced when a larger number of incorrect initial associations were made. The finding is consistent with other recent research suggesting that adults benefit from the spacing of word learning opportunities over time, which also creates initial difficulties in processing (Vlach & Sandhofer, 2014).

The inverse effect of initial accuracy not only suggests parallels between word learning and other domains of cognitive functioning, such as syllogistic reasoning, categorization, and memory (e.g., Alter et al., 2007; Mueller & Oppenheimer, 2014; Oppenheimer, 2008; Oppenheimer & Frank, 2007), but also highlights the possibility that at least under certain circumstances cross-situational learning may be sensitive to the error signal produced when learners' internally generated expectations are no longer supported by experience. The effect, however, remains to be understood. In perhaps its most straightforward (associationist) interpretation, it involves learners keeping track of both strong and weak word-picture associations from the first presentation block, noticing subsequently the lack of support for the initially incorrect dominant associations along with the support for the alternative weaker associations, and then correcting, for example, by exchanging the status of the two associations.¹ Similar to current associative accounts of cross-situational word learning (Kachergis, Yu, & Shiffrin, 2014; Smith & Yu, 2013; Smith et al., 2014; Vlach & Johnson, 2013), this interpretation highlights the role of memory and attention. It also extends these accounts by hypothesizing that the accumulation of evidence entails noticing and correcting error (though not necessarily consciously). The proposal is consistent with a long-standing theoretical interpretation of associative learning as involving the minimization of error from prediction (Rescorla & Wagner, 1972) and evidence for that interpretation from animal and human research (e.g., Fletcher et al., 2001; Luque, López, Marco-Pallares, Càmara, & Rodríguez-Fornells, 2012).

The potential presence of error-related processes in adult cross-situational word learning further highlights the need for developmental research to understand how these processes function over development. The capacity for error detection is present in infancy (e.g., Berger, Tzur, & Posner, 2006; Casey et al., 1997) but undergoes major developments that extend into adolescence. These developments have been associated with changes in processing speed, strategy use and efficiency, individual difference variables, and genetically controlled changes to the anterior cingulate cortex (ACC), the main brain area associated with error processing, and the cholinergic and dopaminergic signaling systems involved in reinforcement learning (Checa, Castellanos, Abundis-Gutiérrez, & Rosario Rueda, 2014; Chevalier, Huber, Wiebe, & Espy, 2013; Grammer, Carrasco, Gehring, & Morrison, 2014; Posner, Rothbart, Sheese, & Tang, 2007; Posner, Rothbart,

Sheese, & Voelker, 2012; Zelazo, Carlson, & Kesek, 2008). Although most of these underlying changes appear to be quantitative, there are suggestions for qualitative shifts (e.g., in cognitive control neuromodulators and the ACC around age 4, Posner et al., 2012; Ramscar & Gitcho, 2007; and in executive control strategies around age 10, Chevalier et al., 2013). Thus, it remains unclear whether children would exhibit the same effects with respect to error in cross-situational learning as adults. Furthermore, should there be differences, it remains unclear whether they would be quantitative or qualitative.

1.4. Research outline

The present experiment was designed to chart the developmental trajectory of the role of initial accuracy in word learning by including 4-year-olds, 10-year-olds, and young adults in the same study. Participants were randomly assigned to two conditions. In the familiarization phase, they consecutively saw ten picture-pseudoword pairs. In the High initial accuracy (IA) condition, six of these pairs were part of the to-be-learned language (accurate items). In the other four pairs, the words and pictures were mismatched (inaccurate items). In the Low IA condition, four pairs in the familiarization phase were part of the to-be-learned language and six were mismatched. The random assignment of participants to condition is an important feature of this experiment as a potential issue for Fitneva and Christiansen's (2011) findings is that participants' assignment to groups was determined post hoc from participants' eye movements in the first block.

The design also allowed us to address another question that was overlooked by Fitneva and Christiansen (2011): Does the inverse effect of initial accuracy result from the properties of individual words in each condition (item-driven learning) or does it also reflect influence among the words (system-driven learning)? To illustrate, the cognitive effort explanation of the effect could operate in two ways. In what we call item-driven learning, the effect could be simply a matter of numbers. If initially inaccurate items are learned better than initially accurate ones due to eliciting more cognitive effort, the greater proportion of the former in the Low IA condition is sufficient to explain learners' better performance in this condition. Alternatively, in what we call system-driven learning, the effect emerges from the cognitive resources recruited by initially inaccurate items affecting initially accurate items as well. Here, the inverse effect results from all items in the Low IA condition being learned better than the corresponding items in the High IA condition, and it is not necessary that initially inaccurate items are learned better than initially accurate ones. Importantly, the question about the contribution of item-driven and system-driven processes can be posed also if a direct effect of initial accuracy is observed.

This secondary focus of our research has potential implications for understanding language learning in general and cross-situational learning in particular. A large part of word learning research pursues its goals by employing few and unrelated items. Yet a number of investigations have suggested that properties of the system, such as item presentation order, matter (e.g., Elman, 1993; Newport, 1990). The cross-situational learning paradigm itself capitalizes on the insight that learners can exploit what they have learned, or

partially learned, to disambiguate novel learning situations (e.g., Yu & Smith, 2007; Yurovsky, Yu, & Smith, 2013; Yurovsky et al., 2014). In these studies, however, knowledge, is at least partially “on the right path,” or correct given the target language. In contrast, we squarely placed participants “on the wrong path” for some of the items, thus exploring an alternative way in which items may influence each other in cross-situational learning.

2. Method

2.1. Participants

The participants in the study were 35 4-year-olds ($M = 4$ years 8 months, range 4;1 to 5;0, 22 girls), 31 10-year-olds ($M = 10$ years 7 months, range 9;10 to 11;0, 11 girls), and 37 undergraduate students ($M = 18.6$ years, range 17–22 years, 33 women). Additional four 4-year-olds were tested but not included in the analyses because of technical problems (1), lack of understanding (English was not a native language, 1), and distraction during test (2). Twelve adults were also tested but not included in the analyses because of extensive non-English language background (e.g., other native language(s) and language spoken at home, 9) or lack of compliance with the instructions (3). Given the demographics of the student population, it was impractical to use language background as a selection criterion at time of recruitment. Children received a small gift and a certificate while students received course credit for their participation.

2.2. Materials

Ten monosyllabic pseudowords, all phonologically valid in English, were used in the study. The pseudowords were recorded by a trained female native English speaker. Ten pictures of different kinds of orchids served as referents. Our goal was to use real, clearly nameable but unfamiliar objects of roughly equal appeal and interest. (The only label undergraduates provided for the pictures in piloting were “flower” and “orchid.”) The materials for children included two additional orchid pictures and two additional pseudowords (used to explain the learning task) and four pictures of common objects (doll, fish, pear, and peach) and the corresponding words (used to demonstrate the testing procedure).

2.3. Design

The study had a familiarization—learning—test structure. The pairing of pictures and pseudowords was randomized across participants thus creating a unique target language for each learner. The *familiarization phase* served to introduce associations between the pictures and pseudowords. In the High IA condition 60% of these associations were accurate (i.e., corresponded to the target language); in the Low IA condition 40% were

accurate. The percentages were chosen to be close to Fitneva and Christiansen's (2011) High and Low IA groups. The inaccurate picture-pseudoword associations were created by scrambling the pseudoword-to-picture assignments.

The *learning phase* consisted of three blocks of five trials each. Each learning trial combined two picture-pseudoword pairs of the target language. Each picture-pseudoword pair appeared three times, once in each block, with picture location (left or right) and order of its label (first or second) counterbalanced across presentations. Two picture-pseudoword pairs appeared together only once. As they were randomly paired, however, an initially inaccurate item could appear with an initially accurate or an initially inaccurate one.² Thus, the block in which an ideal learner could establish the target word-referent associations for initially inaccurate items varied.

In the *test phase*, participants were asked to identify the referent of half of the words of the language. Testing participants on half the items allowed us to avoid presenting images as both targets and foils, thus ensuring the independence of each test trial. Two test lists of five items each were generated. For about half of the participants, the pictures on one of the lists served as targets, and pictures on the other list served as foils; for the other half, targets and foils were exchanged. There were three blocks of test trials for altogether 15 test trials. The location of the target and foil as well as trial order were counterbalanced across block.

2.4. Procedure

Participants were tested individually and were randomly assigned to condition and test list. The audio stream was presented over headphones. The experiment lasted about 5 min and the experimenter monitored participants for their attention to the task through the entire study.

For the familiarization phase, participants were told that they would be first shown the pictures they would be learning names for and hear the words they would have to learn. Adults were also told that this was part of a computer calibration procedure. On each familiarization trial, participants saw a fixation point in the middle of the screen for 250 ms. Four hundred milliseconds later, a single image appeared on the screen and stayed there for 3 s. One second after its appearance a pseudoword was played.

For the learning phase, participants were instructed that their task was to figure out which image corresponded to which picture. Each trial began with a fixation point in the middle of the screen for 250 ms. Four hundred milliseconds later, the two images were presented simultaneously on the screen and stayed there for 4,500 ms. The first pseudoword was presented 1,000 ms and the second 2,750 ms after the appearance of the images on the screen.

For the test trials, participants were told that they had to identify which of two images corresponded to a pseudoword. They were instructed to place their index fingers on the computer keyboard and to use the 1 and 2 keys to select the left and right image, respectively. The keys were marked with stickers for the children, and the keyboard was positioned so that the response keys were directly in front of the participant. On each trial,

after a 250 ms fixation point and a 400 ms blank screen, two images were presented for 750 ms before a pseudoword was played. After participants identified a referent for the pseudoword, a border was placed around the selected image for 500 ms.

There were three main differences between the procedures with children and adults. First, adults read the instructions for each phase while children heard pre-recorded instructions and verbal explanations that used simplified language. Second, for children, the study began with the experimenter giving examples of the objects they had to learn labels for and the pseudowords. These pictures and pseudowords were not used in the study. Third, following this demonstration, the experimenter explained how to use the keyboard to respond at test. The explanation used English words and images (pear-fish, doll-peach). It was also given to adults but no demonstrations were included.

3. Results

3.1. Main analysis

We used generalized estimating equations (GEEs) to model the variability in the data as this technique does not assume homoscedasticity or normality, which is violated by binary repeated measures data. We report Wald χ^2 tests of significance for the main effects and interactions (for previous use of this technique, see Cimpian & Scott, 2012; Fitneva, 2010). The likelihood of correctly identifying the referent of a word was modeled as a function of age (4, 10, and adult), familiarization condition (High and Low IA), and item category (initially accurate or inaccurate). In this model, familiarization condition effects correspond to system-level effects and item category effects correspond to item level effects. Given that the test lists were the same for all participants while word-referent pairing was individually randomized, test lists were not matched for the frequency of initially accurate and inaccurate items either within or between condition, nor were the item characteristics of target–foil pairs controlled. GEE techniques are robust with respect to this data structure and more generally with respect to missing data (Twisk & de Vente, 2002). As the point at which the target pseudoword-picture associations could be made for initially inaccurate items was not controlled and the frequencies of these points in a learning block varied between participants (in addition to varying between condition), we did not include this information as a variable in the analyses.

Fig. 1 shows the proportion of accurate responses at test by age and condition. Accuracy significantly improved with age, Wald $\chi^2(2) = 41.026, p < .001$. Overall, the effect of condition was not significant, Wald $\chi^2(1) = 2.092, p = .148$, and participants performed better on initially accurate than initially inaccurate items, Wald $\chi^2(1) = 4.966, p = .026$. However, as Fig. 1 suggests, there was a significant age by condition interaction, Wald $\chi^2(2) = 11.790, p = .003$. There was also a significant age by item category interaction, Wald $\chi^2(2) = 11.790, p = .003$. The two-way interaction between condition and item category and the three-way interaction between age, condition, and item category were not significant, Wald $\chi^2(1) = .001, p = .974$ and Wald $\chi^2(2) = .740, p = .691$, respectively.

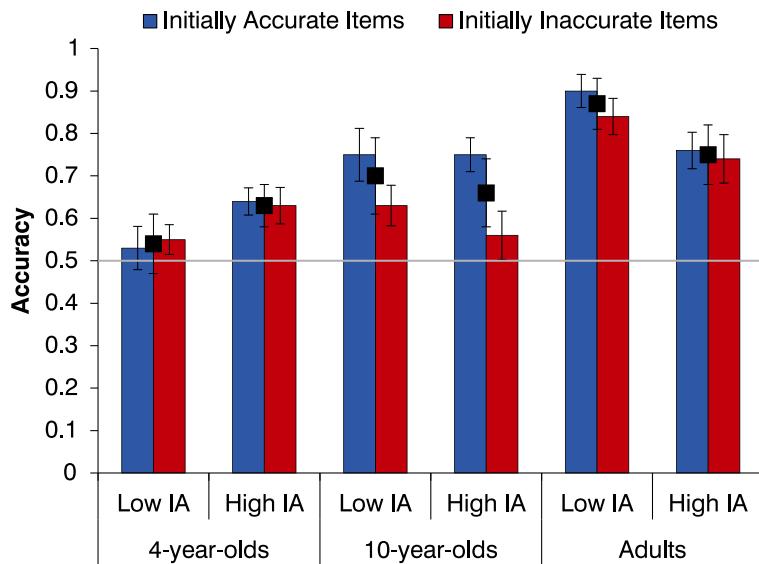


Fig. 1. Mean accuracy (± 1 SE) at test by age, familiarization condition (High IA and Low IA), and item category (initially accurate and initially inaccurate). The black squares show the condition averages with 95% CI.

To assess the effects of condition and item category in each age group, we analyzed the data of each age group separately. Replicating Fitneva and Christiansen's (2011) results, adults in the present study were significantly more likely to correctly identify the referent of a word in the Low than the High IA condition, Wald $\chi^2(1) = 7.313$, $p = .007$. The effect of item category was not significant, Wald $\chi^2(1) = .855$, $p = .355$, nor was the interaction effect between condition and item category, Wald $\chi^2(1) = .397$, $p = .528$.

Ten-year-olds responded similarly in the two conditions, Wald $\chi^2(1) = .542$, $p = .462$. They were significantly more likely to correctly identify the referent of a word if word and referent were paired correctly from the start, Wald $\chi^2(1) = 8.094$, $p = .004$. The interaction between condition and item category was not significant, Wald $\chi^2(1) = .316$, $p = .574$.

Four-year-olds' performance, as adults', was affected by condition. However, initial accuracy had a direct rather than an inverse effect on their performance: They were significantly more likely to correctly identify the referent of a word in the High than the Low IA condition, Wald $\chi^2(1) = 3.963$, $p = .046$. The effect of item category and the interaction effect were not significant, Wald $\chi^2(1) = .027$, $p = .87$ and Wald $\chi^2(1) = .202$, $p = .653$, respectively.

We examined the 95% confidence intervals to assess if performance was different from chance for each age group in each condition (see Fig. 1). With the exception of 4-year-olds in the Low IA condition all groups performed significantly above chance (p 's $< .01$). Further supplemental analyses examined if the results were affected by (a) whether the foil was initially accurate or inaccurate; (b) whether a target-foil pair at test had appeared in the learning phase; and (c) gender (for 4- and 10-year-olds only as there were only four males in the adult group). The variables were included separately in the analyses of

each age group's data because of power considerations. The only significant effect was an interaction between target and foil initial accuracy for 4-year-olds, Wald $\chi^2(1) = 5.783, p = .016$. Four-year-olds performed worse when target and foil were the same category (i.e., both initially inaccurate or both initially accurate, $M_s = .55$ and $.52$, respectively) than when they were different categories ($M = .65$ in both cases). The effect of condition in this analysis was slightly reduced, Wald $\chi^2(1) = 3.399, p = .065$. Because there were no other detectable effects of foil initial accuracy and we had no predictions concerning this factor, we do not discuss it further.

3.2. Performance on initially accurate items

The lack of an effect of item category on adults, and 4-year-olds' performance indicates system-driven learning outcomes for these age groups. In contrast, the item category effect on 10-year-olds' performance suggests item-driven learning outcomes. To provide another test of the contribution of system- and item-level processes, we examined separately learners' performance on the initially accurate items. Participants in both familiarization conditions saw these items the same number of times before test (four, including the familiarization trial). Thus, if learning outcomes are item-driven, performance on these items should not differ between condition. However, if learning outcomes are system-driven, then we may see differences that mirror participants' overall performance. Consistent with system-driven effects on learning, adults performed better on initially accurate items in the Low IA condition than in the High IA condition, Wald $\chi^2(1) = 4.442, p = .035$.

In contrast to adults, 4-year-olds performed better on initially accurate items in the High than the Low IA condition, Wald $\chi^2(1) = 3.205, p = .073$. Given the direct effect of initial accuracy on learning in this group, a stringent test of system-driven effects requires examining performance on initially inaccurate items. However, the design was not optimized for such a test. Nevertheless, the difference in performance on initially accurate items in the two conditions is indicative of system-driven learning. Consistent with the item-level interpretation of 10-year-olds' performance, their performance on initially accurate items did not differ between conditions, Wald $\chi^2(1) = .012, p = .912$.

3.3. Analysis of learning using an alternative criterion

We next examined the data setting the criterion for learning to accurately responding on all three trials involving an item. According to this criterion, on average 4-year-olds learned one word, 10-year-olds two words, and adults three words out of the five they were tested on. Fig. 2 shows the proportion of initially accurate and initially inaccurate items that were learned as a function of age and condition. Fig. 2 confirms the trends from the previous analysis. For ease of interpretation, here we report the results of a repeated-measures ANOVA with age and condition as between-subject variables and item initial accuracy as a within-subject variable. (The same results were obtained when the data were analyzed using GEE.) Seven participants (two 4-year-olds, three 10-year-olds,

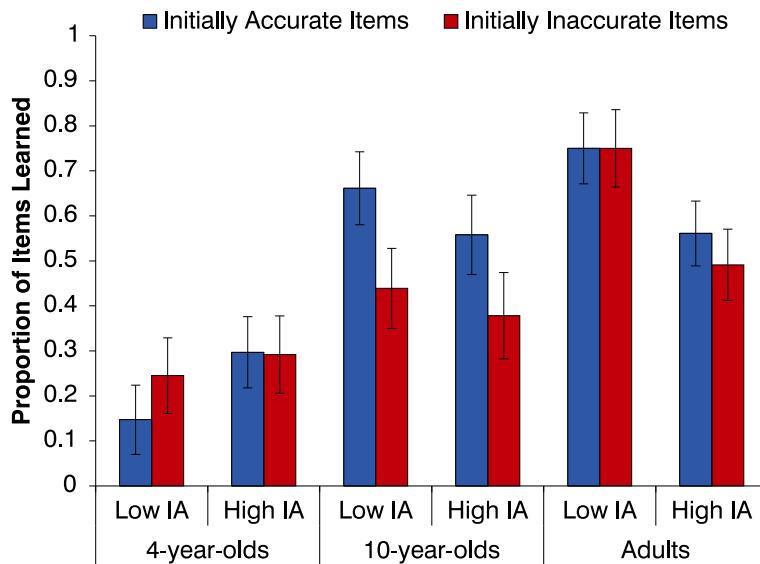


Fig. 2. Average proportion of items learned (three correct answers on three trials), by age, familiarization condition (High IA and Low IA), and item category (initially accurate and initially inaccurate). Error bars represent ± 1 SE.

and two adults) were excluded from this analysis because all items they were tested on were either initially accurate or initially inaccurate.

The analysis showed that with age, participants learned more items to criterion, $F(2, 90) = 23.343, p < .001, \eta_p^2 = .34$. Furthermore, there was a significant age by condition interaction, $F(2, 90) = 3.814, p = .026, \eta_p^2 = .08$. No other effects were significant. Follow-up analyses for each age group showed that adults learned more words in the Low than the High IA condition, $F(1, 33) = 6.714, p = .014, \eta_p^2 = .17$. Four-year-olds learned more words in the High than the Low IA condition but the difference was not significant, $F(1, 31) = 2.085, p = .159, \eta_p^2 = .06$. The effect of item category and the interaction effect between condition and item category were not significant for either group. The only significant effect for 10-year-olds was that of item category, $F(1, 26) = 5.306, p = .03, \eta_p^2 = .17$. They learned more initially accurate items to criterion than initially inaccurate ones. As Fig. 2 shows, however, like adults, this group showed a trend to learn more items to criterion in the Low than High IA condition.

4. Discussion

The goal of the present research was to contribute to understanding cross-situational word learning skills and their development between preschool and adulthood. Using an experimental test, we replicated the finding that, for adults, greater initial inaccuracy in word-referent mappings is associated with better learning (Fitneva & Christiansen, 2011).

We extended this result in two directions. First, we compared adults' performance with that of preschoolers and elementary school children, discovering substantial developmental differences. In particular, preschoolers showed the more intuitive direct effect of initial accuracy: they learned better in the High than the Low initial accuracy condition. Second, we examined whether the inverse effect of initial accuracy was related only to individual item properties or whether items influenced each other's learning. We explored this issue in the child data as well, discovering again considerable age-related changes. Both adults and 4-year-olds learned the words as a set with participants in the Low and High IA conditions, respectively, performing better on all items relative to same-age participants in the other condition. In contrast, 10-year-olds performed better on initially accurate than initially inaccurate items and were unaffected by condition. In the rest of the discussion, we first focus on the inverse effect of accuracy and then on the developmental trends revealed by the data.

4.1. The inverse effect of initial accuracy in adults

The replication of the inverse effect of initial accuracy has important theoretical implications for models of cross-situational learning and for understanding its development. In particular, the findings confirm that erroneous initial associations may result in learning benefits for adults. Although present computational models of cross-situational word learning have considerable coverage when it comes to empirical word learning data, including the existence of erroneous associations (e.g., McMurray et al., 2012; Siskind, 1996; Trueswell et al., 2013; Yu & Smith, 2012; Yurovsky et al., 2014), it is not clear how they can deal with the present results. Specifically, a challenge for current models—whether associative or not—is to take the trial-by-trial history of the learning process into account to capture the beneficial effect of making initial incorrect mappings on subsequent learning. Moreover, a further developmental factor may need to be included to explain the changing role of initial accuracy across development and the changes between system-driven and item-driven word learning.

While contradicting most views on the role of initial accuracy in word learning, our evidence for an inverse effect of initial accuracy in adult learning is consistent with the growing evidence for "desirable difficulties" in learning (Bjork, 1994): A learner's experience of disfluency appears to confer advantages, probably through encouraging more systematic processing (e.g., Alter et al., 2007; Mueller & Oppenheimer, 2014; Oppenheimer & Frank, 2007). In cross-situational learning, such experience may support learning by creating conditions that promote attention, the encoding and storage of information, or cross-modal integration. In general, the inverse effect of initial accuracy invites the integration of current models of cross-situational learning with expectation-based approaches to learning in which prediction-error minimization is the driving force behind learning (e.g., Clark, 2013; Ramscar & Yarlett, 2007; Rescorla & Wagner, 1972).

The present findings also showed that the inverse effect of accuracy in adult cross-situational word learning is not due only to how initially inaccurate items themselves are processed but to how the to-be-learned system is affected by the presence of associations with

such items. Initially inaccurate items were not learned better. Instead, all items, whether initially inaccurate or accurate, were learned better when they were part of a language with more initially inaccurate items. Both computational and experimental studies have suggested that learning is affected by the relations between the items being learned (e.g., Elman, 1993; Newport, 1990). Cross-situational learning in particular is associated with the exploitation of partial knowledge of some word-referent associations to bootstrap the learning of other associations (Yurovsky, Yu, et al., 2013; Yurovsky et al., 2014). Our results suggest that the processing of individual words affects the processing and learning of other words *beyond* the point of the identification of their meaning. Indeed, initially accurate mappings were established from the start and they should have facilitated to a greater extent the discovery of other mappings in the High than the Low initial accuracy condition. Thus, the inverse effect of initial accuracy points to alternative reasons—such as benefits from cognitive disfluency—for why the outcome of learning a set of words that are related through cross-situational learning is different from learning these words individually.

We offer no direct evidence as to the mechanism behind the inverse effect of accuracy in adults. In this regard it would be appropriate to supplement behavioral data with neuroimaging techniques that could shed further light on the neurocognitive correlates of error processing in cross-situational word learning. Our findings raise a question in particular about the involvement of cognitive control networks. Neuroimaging results suggest that the ACC is activated by processing disfluency (Boksem et al., 2005), which we hypothesize our adult participants have experienced after making an incorrect initial mapping. Part of the medial frontal lobe and connected to the prefrontal and parietal cortices, the ACC has been implicated in the prediction and detection of errors in behavior (Carter et al., 1998; Modirrousta & Fellows, 2008; see Rushworth, Behrens, Rudebeck, & Walton, 2007 for review), especially under difficult task conditions, as in early learning (Allman, Hakeem, Erwin, Nimchinsky, & Hof, 2001). Studies also indicate that the ACC plays an important role in exploratory information-seeking behavior in indeterminate learning situations (Forstmann, Brass, Koch, & von Cramon, 2006; Walton, Devlin, & Rushworth, 2004). Within the context of the inverse effect of initial accuracy, these neuroscientific findings point to the ACC networks as an area of interest for investigating the neural bases of cross-situational learning.

Another limitation of our findings is that they are constrained to the learning of a small set of words. Further research is required to understand whether the inverse effect of initial accuracy scales up to learning larger sets of words and sets with complex relations among the items like the human language. It is also incumbent to ask, if the experience of error creates desirable difficulties for the word learner, what the boundaries of desirable difficulty are (cf. Yue, Castel, & Bjork, 2012). This issue is particularly important given the developmental differences in performance, to which we turn next.

4.2. Developmental change

Hitherto, work on cross-situational learning has underscored the existence of this powerful statistical learning mechanism over the lifespan (McMurray et al., 2012; Medina

et al., 2011; Smith & Yu, 2008; Smith et al., 2014; Suanda, Mugwanya, & Namy, 2014; Yu & Smith, 2007, 2012). Correspondingly, with the exception of 4-year-olds in the Low IA condition, we found that participants of all ages were able to exploit the statistical information in the input to find and learn the word-referent mappings, even though they had been misled about some of them. However, we also observed two important developmental changes. First, when children and adults were compared under the same conditions, there was a clear improvement of learning with age. This change is not surprising given recent evidence about the role of attention and memory in both adult and infant cross-situational word learning (e.g., Vlach & Johnson, 2013; Vlach & Sandhofer, 2014), and it appears to be gradual. The second change, however, is more striking. Its two faces are the transition from a direct to an inverse effect of initial accuracy on learning and the transition from system-based to item-based to system-based learning between the ages of 4, 10, and 18. These transitions, which we will treat together, suggest qualitative changes in cross-situational learning over the lifespan. In the rest of this section, we focus first on two concerns about these data, then on two other explanatory avenues, and finally on the practical and theoretical implications of this finding.

One concern about the age-related changes in the effect of initial accuracy is that they stem from differences in the efficacy of our manipulation across groups. The inverse effect of initial accuracy assumes binding of auditory and visual information in the familiarization phase. We cannot ascertain that this binding happened, even though the experimenter monitored participants' attention to the task throughout the study (cf. Smith & Yu, 2013, for a discussion of this issue in infant research). Mitigating this concern is the fact that within each age group there was either an effect of condition or an effect of initial item accuracy suggesting that the familiarization phase affected all groups. Even though there may still be encoding differences between the groups, learners appear to have associated the concurrent visual and auditory input. Clearly, deeper encoding of the word-referent associations during the familiarization phase may result in different levels of disfluency, so the possibility of encoding differences is important to address in future research. Our point is that the premise for studying the inverse effect of initial accuracy—word-referent binding in the familiarization phase—appears to have been met.

A related second concern is that the task demands differed for the three age groups. Learning ten associations is almost certainly not as difficult for adults as for 4-year-olds. If the attention-memory system is already taxed, learners may not be able to correct the initially incorrect associations, benefit from that correction, or aggregate information across trials. Thus, it is conceivable that cognitive limitations contribute to the observed developmental trends beyond the overall age-related improvement in learning. Before any firm conclusions are drawn about developmental differences in the inverse effect of initial accuracy, children's performance on smaller learning sets should be examined.

Next we discuss two other directions in which research can seek explanation for the qualitative changes in cross-situational learning captured by the present findings. Essentially, we offer speculations about the nature of the developmental factor that, as we pointed out, current computational models (e.g., McMurray et al., 2012; Siskind, 1996; Yu & Smith, 2012) may need to include in order to explain the changing role of initial

mappings across development and the changes between system-driven and item driven lexical learning.

The first direction concerns neurodevelopment. As we suggested, the inverse effect of accuracy implicates cognitive control mechanisms in cross-situational learning. These mechanisms, including the frontal cortex and in particular the ACC, undergo substantial changes in their anatomy and functional connectivity through childhood and adolescence (Casey et al., 1997; Davies, Segalowitz, & Gavin, 2004; Posner et al., 2007, 2012; Van Meele, Heslenfeld, Rommelse, Oosterlaan, & Sergeant, 2012), with some researchers explicitly arguing that the maturation of the ACC may underlie many of the differences in child and adult language learning (Ramscar & Gitcho, 2007). The ability to take advantage of the implicit feedback on initially incorrect mappings may require a fully developed ACC network and underlie the relatively late emergence of the inverse effect of initial accuracy.

The second direction in which research can seek explanation for the markedly non-linear developmental changes in the present study involves the ecology of word learning over development. A number of scholars have remarked that the context of language learning changes dramatically from early to middle childhood and adulthood (e.g., Anglin, 1993; Karmiloff-Smith, 1986; Nippold, 1998, 2002). These changes concern the background linguistic knowledge of the learner, the role of others (e.g., parents, teachers), and the structure of the input. To mention but a few, infants and young children have limited knowledge of language and the world. Ostensive teaching of words is much more common to them than school-age children and adolescents (Dockrell & Messer, 2004). In addition, several studies now suggest that parents of young children create mostly unambiguous naming episodes (Frank, Tenenbaum, & Fernald, 2013; Yurovsky, Smith, & Yu, 2013). In comparison, 10-year-olds have extensive linguistic and world knowledge. At this age, formal schooling promotes a stance to language as an object of study. In part as a result of literacy skill development, in this period there is rapid vocabulary growth for more abstract and morphologically complex words and growth in metalinguistic awareness (Anglin, 1993; Ravid & Tolchinsky, 2002). Distinct from both 4- and 10-year-olds, young adults are mostly autonomous word learners. They employ their vast knowledge and reasoning to interpret novel words, which are predominantly slang and jargon (Nippold, 1998). In sum, the context, and thus the problem of word learning changes profoundly over development. Unfortunately, the “task analysis” in language development research has almost exclusively focused on infants (cf. the poverty of the stimulus argument in Chomsky, 1965) and thus the role of these changes has not been systematically examined.

Looking at the word learning task from the point of view of and within the socio-cultural context faced by the learner at different ages, however, may help find an explanation of the non-linear developmental changes observed in the present study. In particular, it may not be (only) cognitive and neuromaturational changes that underlie the observed differences but (also) *how* learners deploy their existing capacities. If infants and young children have relatively clean data, they can just not attend to errors when detected—not because of lack of ability but because of motivation or the perceived importance of such

events. Furthermore, the transition from system-based to item-based to system-based learning may relate to learning behaviors fostered by the academic system rather than learners' abilities. Thus, further research integrating socio-cultural variables may benefit our understanding of developmental changes in learning.

The non-linear qualitative changes observed in the effect of initial accuracy over development have important methodological and theoretical implications. Methodologically, they question the premise of much research using adults to study language learning, namely that adult learning presents a faithful model of learning at other ages (e.g., the "human simulation paradigm" employed by Gillette, Gleitman, Gleitman, & Lederer, 1999). Theoretically, in addition to calling for a revision of existent computational models, our findings suggest that development in cross-situational learning does not result exclusively from quantitative increase in the efficiency of associated memory and attention processes; it appears to be also driven by qualitative changes associated with the role of error over age. Because of the 6- to 8-year gap between the groups, it remains unclear whether these qualitative changes reflect clear-cut discontinuities in development or, alternatively, gradual changes in the environment and in the multiple cognitive networks that support cross-situational learning. Importantly, a qualitative change need not reflect discontinuity. The 10-year-olds' data provide some hints in that direction. Namely, as Figs. 1 and 2 suggest, this group overall performed slightly better in the Low than the High IA condition. This trend appears to foreshadow the inverse effect of initial accuracy in adults.

4.3. Conclusion

The present study suggests that the inverse effect of accuracy may be constrained to the adult language learning system. As much previous research and theory suggest, younger language learners appear to benefit from the accurate initial mapping between words and referents. Even though some differences have been outlined in the language learning skills of preschoolers and adults (e.g., Hudson Kam & Newport, 2005, 2009), we also found substantial differences between 10-year-olds and adults. Middle childhood has received little attention within mainstream language development research, which has traditionally focused on infants and young learners on one hand and adults—as a laboratory model for studying learning—on the other. The apparent effortless gains throughout the lifespan may conceal important changes in the learning mechanism reflecting changes over the course of development both in the learner and the environment.

Acknowledgments

We thank Elaine Choi, Justine Fehr, Justin Goddard, Arielle Mida, and Carolyn Rothenberg for help with data collection, and Peter Elliott for help with data coding. We are also grateful to participating families and students for their cooperation.

Notes

1. A hypothesis-testing account may go as follows. Participants keep track of their commitments from the first presentation block and notice the lack of support for the inaccurate commitments in the second block, which prompts a revision. The outcome of the revision could vary. The revision could lead to an accurate commitment if the word for which an inaccurate commitment has been made is presented along an item that has been mapped correctly. The revision could lead to an accurate or inaccurate commitment if the word is presented with another item that has been mapped incorrectly. If the item is not mapped correctly, the process repeats in the next presentation block. Although error detection plays a prominent role in this model as well, its explanation of the inverse effect of initial accuracy critically depends on assumptions about the trade-off between the problem of late semantic disambiguation (and shorter consolidation experience) and the benefit of cognitive effort generated in the correction process.
2. When two initially incorrect items appeared together, the inaccurate picture—pseudoword associations from the familiarization phase could not be maintained. So the chance of establishing the correct target associations was 50%.

References

- Allman, J. M., Hakeem, A., Erwin, J. M., Nimchinsky, E., & Hof, P. (2001). The anterior cingulate cortex. The evolution of an interface between emotion and cognition. *Annals of the New York Academy of Sciences*, 935, 107–117. doi:10.1111/j.1749-6632.2001.tb03476.x
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136, 569–576. doi:10.1037/0096-3445.136.4.569
- Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, 58(10), 1–166. doi:10.2307/1166112
- Berger, A., Tzur, G., & Posner, M. I. (2006). Infant brains detect arithmetic errors. *Proceedings of the National Academy of Sciences*, 103(33), 12649–12653. doi:10.1073/pnas.0605350103
- Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, 126(1), 39–53. doi:10.1016/j.cognition.2012.08.008
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Boksman, K., Theberge, J., Williamson, P., Drost, D. J., Malla, A., & Densmore, M. (2005). A 4.0-T fMRI study of brain connectivity during word fluency in first-episode schizophrenia. *Schizophrenia Research*, 75, 247–263. doi:10.1016/j.schres.2004.09.025
- Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280, 747–749. doi:10.1126/science.280.5364.747
- Casey, B. J., Trainor, R. J., Orendi, J. L., Schubert, A. B., Nystrom, L. E., Giedd, J. N., & ... Rapoport, J. L. (1997). A Developmental functional MRI study of prefrontal activation during performance of a go-no-go task. *Journal of Cognitive Neuroscience*, 9(6), 835–847. doi:10.1162/jocn.1997.9.6.835

- Checa, P., Castellanos, M. C., Abundis-Gutiérrez, A., & Rosario Rueda, M. (2014). Development of neural mechanisms of conflict and error processing during childhood: Implications for self-regulation. *Frontiers in Psychology*, 5, doi:10.3389/fpsyg.2014.00326
- Chevalier, N., Huber, K. L., Wiebe, S. A., & Espy, K. A. (2013). Qualitative change in executive control during childhood and adulthood. *Cognition*, 128(1), 1–12. doi:10.1016/j.cognition.2013.02.012
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Cimpian, A., & Scott, R. M. (2012). Children expect generic knowledge to be widely shared. *Cognition*, 123 (3), 419–433.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03), 181–204. doi:10.1017/S0140525X12000477
- Dautriche, I., & Chemla, E. (2014). Cross-situational word learning in the right situations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(3), 892–903. doi:10.1037/a0035657
- Davies, P. L., Segalowitz, S. J., & Gavin, W. J. (2004). Development of response-monitoring ERPs in 7- to 25-year-olds. *Developmental Neuropsychology*, 25(3), 355–376. doi:10.1207/s15326942dn2503_6
- Diesendruck, G., Carmel, N., & Markson, L. (2010). Children's sensitivity to the conventionality of sources. *Child Development*, 81(2), 652–668. doi:10.1111/j.1467-8624.2009.01421.x
- Dockrell, J. E., & Messer, D. (2004). Lexical acquisition in the early school years. In R. A. Berman (Ed.), *Language development across childhood and adolescence* (vol. 3, pp. 35–52). Amsterdam: John Benjamins.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1), 71–99. doi:10.1016/0010-0277(93)90058-4
- Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, 108(2), 477–499. doi:10.1016/j.cognition.2008.04.002
- Fitneva, S. A. (2010). Children's representation of child and adult knowledge. *Journal of Cognition and Development*, 11(3), 458–484. doi:10.1080/15248371003700023
- Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, 35(2), 367. doi:10.1111/j.1551-6709.2010.01156.x
- Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language*, 36(5), 967–997. doi:10.1017/S0305000908009252
- Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, T. A., Donovan, T., & ... Bullmore, E. T. (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nature Neuroscience*, 4(10), 1043–1048. doi:10.1038/nn733
- Forstmann, B. U., Brass, M., Koch, I., & von Cramon, D. Y. (2006). Voluntary selection of task sets revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, 18, 388–398. doi:10.1162/jocn.2006.18.3.388
- Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, 20(5), 578–585. doi:10.1111/j.1467-9280.2009.02335.x
- Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, 9(1), 1–24. doi:10.1080/15475441.2012.707101
- Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality vs. modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, 19, 117–125. doi:10.1016/j.tics.2014.12.010
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73, 135–176. doi:10.1016/S0010-0277(99)00036-0
- Gori, M., & Tesi, A. (1992). On the problem of local minima in backpropagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(1), 76–86.

- Grammer, J. K., Carrasco, M., Gehring, W. J., & Morrison, F. J. (2014). Age-related changes in error processing in young children: A school-based investigation. *Developmental Cognitive Neuroscience*, 9, 93–105. doi:10.1016/j.dcn.2014.02.001
- Hollich, G. J., Hirsh Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., Hennon, E., & Rocroi, C. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child-Development*, 65(3), 1–123.
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128–157. doi:10.1080/15250000701795598
- Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2), 151–195. doi:10.1207/s15473341lld0102_3
- Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59(1), 30–66. doi:10.1016/j.cogpsych.2009.01.001
- Kachergis, G., Yu, C., & Shiffrin, R. M. (2014). Cross-situational word learning is both implicit and strategic. *Cognitive Science*, 5, 588. doi:10.3389/fpsyg.2014.00588
- Karmiloff-Smith, A. (1986). Some fundamental aspects of language development after age 5. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (2nd ed., pp. 455–474). Cambridge, UK: Cambridge University Press.
- Krogh, L., Vlach, H. A., & Johnson, S. P. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology*, 3, doi:10.3389/fpsyg.2012.00598
- Luque, D., López, F. J., Marco-Pallares, J., Câmara, E., & Rodríguez-Fornells, A. (2012). Feedback-related brain potential activity complies with basic assumptions of associative learning theory. *Journal of Cognitive Neuroscience*, 24(4), 794–808. doi:10.1162/jocn_a_00145
- Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In M. R. Gunnar & M. Maratsos (Eds.), *Modularity and constraints in language and cognition* (pp. 59–101). Hillsdale, NJ: Lawrence Erlbaum.
- McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, 119(4), 831–877. doi:10.1037/a0029872
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, 108(22), 9014–9019. doi:10.1073/pnas.1105040108
- Misyak, J. B., Goldstein, M. H., & Christiansen, M. H. (2012). Statistical-sequential learning in development. In P. Rebuschat & J. Williams (Eds.), *Statistical learning and language acquisition* (pp. 13–54). Berlin: Mouton de Gruyter.
- Modirrousta, M., & Fellows, L. K. (2008). Dorsal medial prefrontal cortex plays a necessary role in rapid error prediction in humans. *Journal of Neuroscience*, 28, 14000–14005. doi:10.1523/JNEUROSCI.4450-08.2008
- Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard advantages of longhand over laptop note taking. *Psychological Science*, 25(6), 1159–1168. doi:10.1177/0956797614524581
- Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, 14(1), Spec Issue, 11–28. doi:10.1207/s15516709cog1401_2
- Nippold, M. A. (1998). *Later language development: The school-age and adolescent years* (2nd ed.). Austin, TX: Pro-Ed.
- Nippold, M. A. (2002). Lexical learning in school-age children, adolescents, and adults: A process where language and literacy converge. *Journal of Child Language*, 29(02), 449–488. doi:10.1017/S0305000902275340
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, 12, 237–241. doi:10.1016/j.tics.2008.02.014

- Oppenheimer, D. M., & Frank, M. F. (2007). A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition*, 106, 1178–1194. doi:10.1016/j.cognition.2007.05.010
- Plagianakos, V. P., Magoulas, G. A., & Vrahatis, M. N. (2001). Supervised training using global search methods. In N. Hadjisavvas & P. M. Pardalos (Eds.), *Advances in convex analysis and global optimization* (pp. 421–432). Dordrecht, the Netherlands: Kluwer.
- Posner, M. I., Rothbart, M. K., Sheese, B. E., & Tang, Y. (2007). The anterior cingulate gyrus and the mechanism of self-regulation. *Cognitive Affective & Behavioral Neuroscience*, 7(4), 391–395. doi:10.3758/CABN.7.4.391
- Posner, M. I., Rothbart, M. K., Sheese, B. E., & Voelker, P. (2012). Control networks and neuromodulators of early development. *Developmental Psychology*, 48(3), 827–835. doi:10.1037/a0025530
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Sciences*, 11(7), 274–279. doi:10.1016/j.tics.2007.05.007
- Ramscar, M., & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science*, 31(6), 927–960. doi:10.1080/03640210701703576
- Ravid, D., & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of Child Language*, 29(02), 417–447. doi:10.1017/S0305000902005111
- Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, 29, 819–865. doi:10.1207/s15516709cog0000_31
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton Century Crofts.
- Rushworth, M. F. S., Behrens, T. E. J., Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends in Cognitive Sciences*, 11, 168–176.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. doi:10.1126/science.274.5294.1926
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8(2), 101–105. doi:10.1111/j.1467-9280.1997.tb00690.x
- Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development*, 69(1), 94–104. doi:10.2307/1132073
- Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61, 39–91. doi:10.1016/S0010-0277(96)00728-7
- Smith, L. B. (2000). How to learn words: An associative crane. In R. Golinkoff & K. Hirsh-Pasek (Eds.), *Breaking the word learning barrier* (pp. 51–80). Oxford, UK: Oxford University Press.
- Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word-referent learning. *Trends in Cognitive Sciences*, 18(5), 251–258. doi:10.1016/j.tics.2014.02.007
- Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106, 333–338.
- Smith, L. B., & Yu, C. (2013). Visual attention Is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development*, 9(1), 25–49. doi:10.1080/15475441.2012.707104
- Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, 126, 395–411. doi:10.1016/j.jecp.2014.06.003
- Toro, J., Sinnott, S., Soto-Faraco, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97, B25–B34. doi:10.1016/j.cognition.2005.01.006
- Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, 66(1), 126–156. doi:10.1016/j.cogpsych.2012.10.001

- Twisk, J., & de Vente, W. (2002). Attrition in longitudinal studies. How to deal with missing data. *Journal of Clinical Epidemiology*, 55(4), 329–337. doi:10.1016/S0895-4356(01)00476-0
- Van Meel, C. S., Heslenfeld, D. J., Rommelse, N. N., Oosterlaan, J., & Sergeant, J. A. (2012). Developmental trajectories of neural mechanisms supporting conflict and error processing in middle childhood. *Developmental Neuropsychology*, 37(4), 358–378. doi:10.1080/87565641.2011.653062
- Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, 127(3), 375–382. doi:10.1016/j.cognition.2013.02.015
- Vlach, H. A., & Sandhofer, C. M. (2014). Retrieval dynamics and retention in cross-situational statistical word learning. *Cognitive Science*, 38(4), 757–774. doi:10.1111/cogs.12092
- Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, 45(6), 1611–1617. doi:10.1037/a0016134
- Walton, M. E., Devlin, J. T., & Rushworth, M. F. S. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nature Neuroscience*, 7, 1259–1265. doi:10.1038/nn1339
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2), 245–272. doi:10.1037/0033-295X.114.2.245
- Yu, C., & Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414–420. doi:10.1111/j.1467-9280.2007.01915.x
- Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review*, 119(1), 21–39. doi:10.1037/a0026182
- Yue, C. L., Castel, A. D., & Bjork, R. A. (2012). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, 41(2), 229–241. doi:10.3758/s13421-012-0255-8
- Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, 21(1), 1–22. doi:10.3758/s13423-013-0443-y
- Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, 16(6), 959–966. doi:10.1111/desc.12036
- Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, 37(5), 891–921. doi:10.1111/cogs.12035
- Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 553–574). Cambridge, MA: MIT Press.