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

Stanka A. Fitneva
Queen’s University

Morten H. Christiansen
Cornell University

Mailing address:
Department of Psychology
Queen’s University
Kingston, ON K7L 4V1
Canada
Email: fitneva@queensu.ca

Keywords: language development; cross-situational learning; disfluency; statistical learning; prediction-based learning
Author Note

Stanka A. Fitneva, Department of Psychology, Queen’s University and Morten H. Christiansen, Department of Psychology, Cornell University.

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.

Correspondence concerning this article should be addressed to Stanka A. Fitneva, Department of Psychology, Queen’s University, Kingston, Ontario K7L 3N6, Canada. E-mail: fitneva@queensu.ca.
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 suggested that items influenced each other for adults and 4-year-olds but not 10-year-olds. These findings point to a qualitative change in language learning processes over development.
Developmental changes in cross-situational word learning:

the inverse effect if initial accuracy

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, 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 across development other than in closely spaced age groups (Smith & Yu, 2008, 2013; Vlach & Johnson, 2013), which may have obscured in particular the existence of qualitative changes in development.

The goal of the present 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 et al.,
2014; Smith & Yu, 2008; 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. The present study aimed to contribute to establishing the similarities and differences in this type of statistical learning across the ages and thus to 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, namely that 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 has 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).

Cross-situational learning can be described in various ways, the most prominent models advancing hypothesis testing and associative learning. 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 pairing requires correction on a subsequent trial, which is effortful and disruptive, or 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. A few 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 ten words within a cross-situational paradigm. Two pictures and their labels were
presented without there being 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 pairings, 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 e.g., by exchanging the status of
Developmental Changes

the two associations. Similar to current associative accounts of cross-situational word learning (Kachergis, Yu, & Shiffrin, 2014; Smith et al., 2014; Smith & Yu, 2013; 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, Câmara, & 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, 1

1 On an alternative hypothesis-testing account of the inverse effect from Fitneva and Christiansen (2011), participants may 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) created by initial erroneous hypotheses and the benefit of additional cognitive effort created by these hypotheses.
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 is unclear whether children would exhibit the same effects with respect to error in cross-situational learning as adults. Furthermore, should there be differences, it is 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 et al., 2014; Yurovsky, Yu, & Smith, 2013). In these studies 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 to 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 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
objects of roughly equal appeal and interest but still unfamiliar to our participants. (The only label undergraduates provided for the pictures in piloting were “flower” and “orchid.”) Two other pictures of orchids and two other pseudowords were used to explain the learning task to the children. Four pictures of common objects (doll, fish, pear, and peach) and the corresponding words were used to familiarize them with 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 Initial Accuracy (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 assignment among these items.

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 pseudoword order (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. Thus, the block in which the target word-referent associations could be established by an ideal learner 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 randomized 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

---

2 When two initially incorrect items appeared together, the inaccurate picture – pseudoword associations from the familiarization phase could not be maintained and there was 50% chance of establishing the correct target associations.
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 4500 ms. The first pseudoword was presented 1000 ms and the second 2750 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 procedure used 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 their labels. These orchid 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 with English words and images (pear-fish, doll-peach). The explanation was also given to adults but no demonstrations were included.
3. Results

3.1. Main analysis

We used generalized estimating equations (GEE) to model the variability in the data as this technique does not assume normality, which is violated by binary repeated measures data, and also homoscedasticity. We report Wald $\chi^2$ tests of significance for the main effects and interactions (for previous use of this technique, see Cimpian & Scott, 2012; Fitneva, 2010). Given that the test lists were the same for all participants while word-referent pairing was randomized for each participant, 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 pseudowords-pictures associations could be made for initially inaccurate items was not controlled and the frequencies of these events in each learning block varied between participants (in addition to varying between condition), we did not include the point at which an ideal learner could have discovered the correct associations as a variable in the analyses. 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.
Fig. 1 shows the proportion of accurate responses at test by age and condition. As it suggests, 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.

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 they 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 including in the analyses of each group’s data 1) whether the foil was initially accurate or inaccurate; 2) whether a target-foil pair at test had appeared in the learning phase; and 3) 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 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, $Ms = .55$ 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 $\chi^2(1) = 3.399, p = .065$. As 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 and an interaction effect between item category and familiarization condition on adults and 4-year-olds indicate system-driven learning outcomes for these age groups. In contrast, the item category effect on 10-year-olds suggest item-driven
learning outcomes. To provide another test of the contribution of system- and item-level processes, we examined separately the 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 $\chi^2(1) = 4.442, p = .035$.

The same test show that, in contrast to adults, 4-year-olds performed better 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 also consistent with the existence 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, and two adults) were excluded from this analysis because all items they were tested on were either initially accurate or initially inaccurate.

--- Fig. 1 about here ---

The analysis showed that with age, participants learned more items to criterion, $F(2, 90) = 23.343, p < .001, \eta^2 = .34$. Furthermore, there was a significant age by condition interaction, $F(2, 90) = 3.814, p = .026, \eta^2 = .08$. 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, $F(1, 33) = 6.714, p = .014, \eta^2 = .17$. Four-year-olds learned more words in the High than the Low IA condition but the difference was not significant, $F(1, 31) = 2.085, p = .159, \eta^2 = .06$. 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 of item category, $F(1, 26) = 5.306, p = .03, \eta^2 = .17$. 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. Second, we examined whether this result was related only to individual item properties or whether items influenced each other. 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 by adults and its evolution over 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 in order 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 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, it may support learning by creating conditions that promote attention, the encoding and storage of information, or cross-modal integration. In general, this finding 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 to only how initially inaccurate items themselves are processed but to how the to-be-learned system is affected by the presence of inaccurate associations. Initially inaccurate ones were not learned better. But 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). In particular, cross-situational learning involves the exploitation of partial knowledge of some word-referent associations to bootstrap the learning of other associations (Yurovsky et al., 2014; Yurovsky, Yu, et al., 2013). Our results extend these findings to suggest that the processing of individual words affects the processing and learning of other words beyond identifying the target word-referent mappings. Indeed, initially accurate mappings were established from the start. Thus, our results identify another reason 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. Direct examinations of the involvement of neural substrates in cross-situational word-learning have been lacking. Our findings raise a question about the involvement in particular of cognitive control networks. Neuroimaging results suggest that the ACC is activated by processing disfluency (Boksman 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, & 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

Heretofore, work on cross-situational learning has underscored the continuity of the existence of this powerful statistical learning mechanism over the lifespan (McMurray et al., 2012; Medina et al., 2011; Smith et al., 2014; Smith & Yu, 2008; 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 were misled about some of them. However, we also observed two important 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 direct to inverse effect of initial accuracy on learning and the transition from system-based to item-based and again to system-based learning between the ages of 4, 10, and 18. These transitions suggest qualitative changes in cross-situational learning over the lifespan. As understanding the mechanisms behind the developmental trends in the present study was beyond its scope, we treat these transitions together. In the rest of this section, we focus first on two concerns about these data, then on two accounts that can be advanced for the observed qualitative change, and finally on its practical and theoretical implications.

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 are not able to 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.

As we pointed out, current computational models (e.g., McMurray et al., 2012; Siskind, 1996; Yu & Smith, 2012) may need to include a separate developmental factor to explain the changing role of initial mappings across development and the changes between system-driven and item driven lexical learning. Below we speculate about the nature of this factor by discussing two explanations going beyond the issues mentioned above.

One explanation that could be considered for the markedly non-linear developmental changes in our study concerns neurodevelopment. If the analysis we propose of the inverse effect of accuracy is correct, then cognitive control mechanisms are likely implicated 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 Meel, 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.

Another explanation 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 create for them 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 specifically 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 can employ their vast knowledge and reasoning to semantically 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.

The “task analysis” in language development research has almost exclusively focused on infants (cf. the poverty of the stimulus argument in Chomsky, 1965). 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 in our view can contribute to the explanation of the non-linear developmental changes we observe in the effect of the accuracy of initial word-referent mappings. 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 again system-based learning may relate to learning attitudes and 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 in the role of error over age. Because of the 6- to 8-year gap between the groups, it is unclear whether these qualitative changes reflect clear-cut discontinuities in development or, alternatively, gradual changes in the multiple cognitive networks that support cross-situational learning and in the environment. Importantly, a qualitative change need not reflect discontinuity. The 10-year-olds’ data provide some hints in that direction. Namely, as Fig. 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.
References


predicted by formal associative learning theory. *Nature Neuroscience, 4*(10), 1043–1048. doi:10.1038/nn733


Regier, T. (2005). The emergence of words: Attentional learning in form and


Figure Captions

*Figure 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.

*Figure 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.
The Cognitive Science Society, Inc.

Publication Agreement

The undersigned authors agree that the manuscript entitled Developmental Changes in Cross-situational Word Learning: The Inverse Effect of Initial Accuracy

having been accepted for publication in the journal *Cognitive Science*, or *topiCS*, is now the property of the Cognitive Science Society, Inc. Specifically, the author(s) transfer(s) to the Society the exclusive rights comprised in the copyright of said manuscript, except that the author(s) retain(s) the following:

1) All proprietary rights, other than copyright.
2) The right to make oral presentations of the material.
3) The right to use, after publication, all or part of the material in any book authored or edited by the author(s).
4) Authors are entitled to distribute copies of the article for personal use, either on paper or electronically, through their own personal mailing or website, or through the website of an agency by which they are employed, but permission of the Cognitive Science Society is required to reproduce published papers in other sources, including electronic archives.

It is understood, in relation to Item 4, that the author(s) is/are required to obtain permission in writing from the Cognitive Science Society for all or a substantial part of the article to be reproduced for distribution by a commercial publisher or an agency in which the author(s) is/are not employed.

In the case of works prepared by employees or contractors of the U.S. Government, the U.S. Government will retain a non-exclusive, royalty-free license to use any or all of the material for any government purpose.

The author(s) warrant(s) that the manuscript is the original work of the author(s), contains no libelous or unlawful statements, and does not infringe any legal rights of any other party. In the case of an employee work for hire, both the author and a representative authorized to bind the owner of the work have signed below.

Cognitive Science and *topiCS* prefer all co-authors of jointly authored papers to sign the copyright transfer. In the event where only the principle co-author signs, Cognitive Science and *topiCS* will assume that the signing co-author does so with the authorization of the remaining co-author(s).

Intending to be legally bound, and understanding that the Society will grant specific license for republication of this manuscript without charge when agreed to by the authors and upon written request of the prospective licensee, the author(s) have/has subscribed below.

\[Signature\]  
August 7, 2015

Stanka A. Fitneva

Queen's University, Canada

Author's Signature & Date  Printed Name  Institution or Organization

Author's Signature & Date  Printed Name  Institution or Organization

Author's Signature & Date  Printed Name  Institution or Organization

Author's Signature & Date  Printed Name  Institution or Organization

Please sign and upload this form to Editorial Manager with your final revision; or you may email it to the Editorial Office at cogscij@indiana.edu, or topicsj@indiana.edu