

Children automatically abstract categorical regularities during statistical learning

Yaelan Jung  | Dirk B. Walther  | Amy S. Finn

Department of Psychology, University of Toronto, Toronto, ON, Canada

Correspondence

Yaelan Jung, University of Toronto, 100 St. George St, Toronto, ON, Canada M5S 3G3.

Email: yaelanj@princeton.edu

Funding information

Canada Foundation for Innovation, Grant/Award Number: 34947; Canadian Network for Research and Innovation in Machining Technology, Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN-2015-06696 and RGPIN-2016-05; Social Sciences and Humanities Research Council of Canada, Grant/Award Number: 435171493 and 430-2017-01189

Abstract

Statistical learning allows us to discover myriad structures in our environment, which is saturated with information at many different levels—from items to categories. How do children learn different levels of information—about regularities that pertain to items and the categories they come from—and how does this differ from adults? Studies on category learning and memory have suggested that children may be more focused on items than adults. If this is also the case for statistical learning, children may not extract and learn the multi-level regularities that adults can. We report three experiments showing that children and adults extract both item- and category-level regularities in statistical learning. In Experiments 1 and 2, we show that both children and adults can learn structure at the item and category levels when they are measured independently. In Experiment 3, we show that both children and adults learn about categories even when exposure does not require this: both are able to generalize their learning from the item to the category level. Results indicate that statistical learning operates across multi-level structure in children and adults alike, enabling generalization of learning from specific items to exemplars from categories of those items that observers have never seen. Even though children may be more focused on items during other forms of learning, they learn about categories from item-level input during statistical learning.

KEYWORDS

category, children, development, statistical learning, visual

1 | INTRODUCTION

Our world is saturated with regularities at multiple levels, from lower-level sensory features to higher level abstract and semantic relationships that link objects and categories. A longstanding puzzle in psychological science is to understand *how* humans come to know all of these regularities. Statistical learning has been put forth as an answer, as a powerful learning system that can be used to extract a number of different regularities from our highly complex world (Frost et al., 2015). However, given sweeping differences in children's and adults' knowledge of higher level abstract and semantic relationships (Bjorklund, 1988; Gelman & Coley, 1990), it is

likely that children and adults extract *different* regularities from their complex environments. In the current study, we seek to understand which regularities are prioritized in statistical learning and how these differ in children and adults.

1.1 | Statistical learning across multiple levels in adults

Adults are able to extract patterns that are present at multiple levels, including patterns operating at different spatial scales (Fiser & Aslin, 2005; Jun & Chong, 2016), and patterns that occur across both items (Fiser & Aslin, 2001; Turk-Browne et al., 2008) and categories (Brady

& Oliva, 2008; Emberson & Rubinstein, 2016; Otsuka et al., 2013). Indeed, adults have been shown to learn sequences that occur across scene categories (e.g., that a kitchen always comes after a bedroom, but never the same kitchen or bedroom; in Brady & Oliva, 2008) and object-categories (Emberson & Rubinstein, 2016; Otsuka et al., 2013).

One pressing question remains: Is statistical learning constrained to a particular level (or levels) that an observer is exposed to, or can learning be flexibly generalized to more abstract levels? Research has shown that statistical learning can flexibly extend from the level of specific items to the level of basic categories. Brady and Oliva (2008) demonstrated that when exposed to a sequence with specific scene images (e.g., a specific kitchen picture comes after a specific bedroom picture), adult observers are able to show knowledge of the exposed sequences at a semantic level, which was tested using lexical items (e.g., a word “kitchen” comes after a word “bedroom”). These results were also replicated by Otsuka et al. (2013) with line drawings; participants who were exposed to a stream of objects presented as line drawings demonstrated knowledge of the same stream sequence when it was presented using only lexical items. These studies show that statistical learning can generalize beyond exposed items to a known abstract representation of those items—words—but it is as yet unknown if this would also apply to novel visual exemplars of the same category.

1.2 | Statistical learning across multiple levels in children

Whether statistical learning operates at multiple levels is likely to differ in children and adults. Although very young children and infants can extract regularities from their environments (Kirkham et al., 2002; Saffran et al., 1996), their statistical learning differs from adults because of differences in what they know (Finn & Hudson Kam, 2008), the maturity of their brains (Gómez, 2017), and how they interact with their environments (Smith et al., 2018). As yet, however, it is unclear *how* statistical learning differs in children, especially whether children can learn category-level regularities like adults (Brady & Oliva, 2008). Given previous work, there are the following intriguing possibilities.

First, children might be able to extract category-level statistics, as adults have demonstrated. Research has suggested that even very young children have robust knowledge of categories (Gelman & Markman, 1986) and the ability to abstract specific instances to categories (Younger, 1990). Furthermore, research has shown that children are not as good as adults at encoding fine-grained details of specific exemplars within a category and discriminating them in their memory (Keresztes et al., 2017; Ngo et al., 2018), which suggests that children may utilize category-level information (rather than item-specific information) when extracting statistical structure from the environment.

On the other hand, children may be sensitive to item-level features, and therefore learn only the item-level regularities when presented with both. Sloutsky and Fisher (2004) have shown that when children are exposed to different animal pictures and told a

Research Highlights

- Children can learn statistical structure at category level just like adults
- Both children and adults can generalize their learning experience from specific items to multiple novel exemplars that they have never encountered
- When the test requires higher retrieval demands, children, unlike adults, cannot express their learning at the level of category

categorical feature about them (cats but not bears or birds have beta cells), children have *better* memory for the individual pictures than adults who may have only encoded the category-level information. Furthermore, unlike adults who tend to overlook features of an item that are not diagnostic to categories, children remember information regardless of its relation to category membership (Deng & Sloutsky, 2015). Children might therefore be more item-focused (Sloutsky, 2003; Sloutsky et al., 2007) and thus more likely to learn statistical structure at the level of items and *not* the category level.

Finally, children might be able to extract category-level regularities, but only when both item- *and* category-level structure is present in the sequence. When exposed to a stream of images where regularities operate *only* at the category level, observers are presented with noisier structure, where item-level statistics do not provide any meaningful information about the sequence. Thus, although children may be able to track category-level regularities from a stream where both item- and category-level statistics are given, they might not be able to extract category-level statistics when they encounter novel exemplars each time, where no item-level statistics are present.

The current study aims to address all of these possibilities and systematically measure how statistical learning occurs at the item and category levels in children relative to adults. Experiment 1 tests statistical learning at the item level only, Experiment 2 tests statistical learning at the category level only, and Experiment 3, tests category-level learning when participants are exposed to item-level regularities, probing whether knowledge of sequential statistics from specific items can be abstracted to the category level. Altogether, the current study explores what information children can learn from statistical structure operating at the item and/or category levels.

2 | EXPERIMENT 1

2.1 | Methods

2.1.1 | Participants

Thirty adults (18–21 years, mean age = 18.72, SD = 0.85, 26 female) and 30 children (6–9 years, mean age = 7.82, SD = 0.94, 14 female)

participated in Experiment 1. Adult participants were recruited from the University of Toronto and compensated with cash or course credit. Children participated at the Ontario Science Centre and were compensated with a small toy. All participants (and children's parents) consented prior to participation, and the experiment was approved by the Research Ethics Boards of the University of Toronto.

2.1.2 | Stimuli

Five hundred and four images of 12 animal categories were used (42 images for each category) to create all exposure and test stimuli; the categories were comprised of bears, chickens, cows, dogs, dolphins, ducks, elephants, lions, monkeys, seals, sheep, and wolves. Animals from these categories were used to create four triplets (i.e., elephant-chicken-bear; Figure 1a), by randomly assigning three animals to each triplet. Also, four foils were created by randomly selecting an item in each position from different triplets (see Figure 1b). The

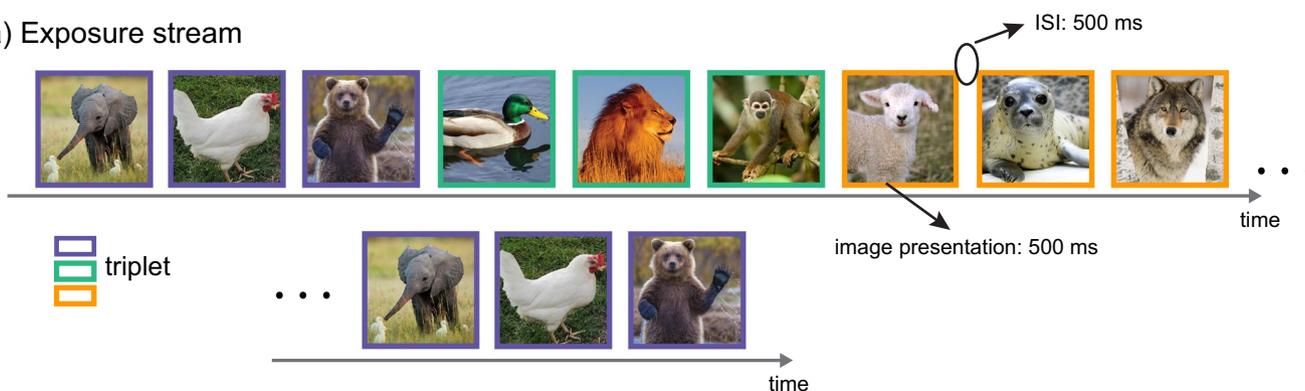
triplet and the foil structure were randomly created for each participant. Moreover, an image from each of the animal categories was randomly selected for each participant. Thus, randomization occurs to both triplet (and foil) structure of the animal categories and to the individual images constructing the triplets for each participant. The stimulus set is available through the Open Science Framework repository (<https://doi.org/10.17605/OSF.IO/U75ZH>).

2.1.3 | Procedure

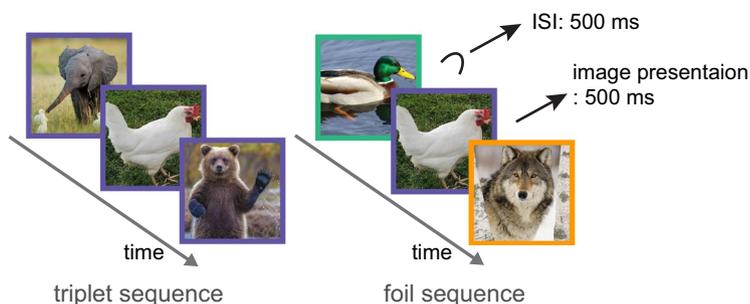
The experiment consisted of the following phases: the initial exposure phase, the two-alternative forced-choice (2AFC) test, and finally a triplet completion test, which always occurred in this order. The triplet completion test was exploratory and is reported in *Exploratory analyses*. Test items and exposure were all presented using Matlab with the Psychophysics toolbox (Brainard, 1997) on a MacBook air screen at a resolution of 1440 × 900 pixels.

Statistical learning of item-level information (Exp 1)

(a) Exposure stream



(b) 2AFC test



(c) 2AFC performance

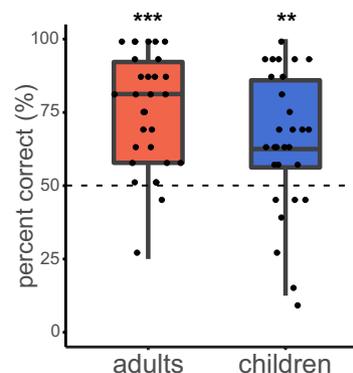


FIGURE 1 Design and results for Experiment 1. (a) Samples from the exposure stream in Experiment 1. Triplets (outlined in unique colors) were each comprised of three animals in the same order and were randomly distributed in the stream. (b) Examples of the test sequences in the 2AFC test. (c) Percentage of triplets chosen as familiar in adults and children. Chance level, indicated by the dashed line, is 50%. Asterisks indicate the performance was significantly higher than chance ($***p < 0.001$, $**p < 0.01$). Each box spans the first quartile to the third quartile, and the central line in the box indicates the median point. Whiskers above and below the box show the location of the minimum and maximum, and dots indicate individual performance

During the exposure phase, participants watched a stream of animals for approximately 5 minutes (Figure 1a). Images were presented one at a time for 500 ms with a 500 ms inter-stimulus interval. Each of the four triplets appeared 24 times, resulting in a total of 288 image presentations. The presentation order of triplets was randomly generated for each participant with the constraint that neither individual triplets nor pairs of triplets could appear twice in a row. Prior to exposure, participants were told to focus on the images, about which they would need to answer questions later. An experimenter stayed in the room with participants at all times to ensure wakeful compliance. After the exposure phase, participants performed two tests of learning, a 2AFC test (Figure 1b) and an exploratory test (Figure 1c).

Participants performed the 2AFC test first. On each test trial, participants viewed two test sequences which had three images each. One of these sequences was a triplet of images that participants saw during the exposure, and the other was a foil constructed with images from three different triplets. The position of each item in the foil sequence was maintained so that it was the same as its assigned position in the triplet structure. As each participant was assigned to a randomly generated triplet structure, the structure of the foils was also randomly generated for each participant. The items in the foils had the same position information as in the triplets. For example, if a dog was the first item in a triplet presented during the exposure, then it was also the first item in the sequence when it appeared in as foil. The first of the two test sequences appeared on the left side of the screen. Following a one-second pause, the second test sequence appeared on the right side of the screen. Similar to the exposure phase, each image in these test sequences was presented one at a time for 500 ms with a 500 ms inter-stimulus interval. After participants viewed both sequences, they were asked to press "1" or "2" to indicate whether the first or the second sequence fits better with what they saw during exposure.

The experimenter instructed participants verbally with age-appropriate language (e.g., "You will see three pictures at each side of the screen. Tell me which group feels like something you saw before."). To confirm that all participants (especially children) understood the task, three practice trials using numbers (i.e., 1-2-3, vs. 2-1-3; and 4-5-6 vs. 6-4-5) and letters (a-b-c vs. b-a-c) were performed before the experimental trials. During the practice, participants were asked which of the sequences was in the correct order. To introduce the phrase "being in the correct order" to the children, the experimenter put alphabet cards in the correct (a-b-c) and incorrect (b-c-a) orders one at a time. The experimenter read one card at a time (a, b, c) before asking participants whether this is right or silly. The experimenter confirmed the participants' answer using a phrase such as: "Yes. This is right because b comes after a, c comes after b. When it is right, we say this is in order!" After learning this phrase with the cards, participants performed the practice trials on the computer. If participants did not respond correctly to all of three practice trials, the experiment did not continue. During the test, each of the four triplet sequences were tested four times, exhaustively paired with the four foils, which were thus also tested four times each. There was an equal chance of the triplet or the foil being

presented first. Participants had the option to view each trial one more time before making a response. After completing the 2AFC test, participants also performed the exploratory test, and the entire experiment took about 20 minutes. The procedure and the results from the exploratory test will be presented in Section 5.

2.1.4 | Data analyses

One-tailed, one sample *t*-tests were used to determine if 2AFC test performance was greater than chance (50%). Independent sample *t*-tests (two-tailed) were conducted to compare the test performance between age groups. Analyses of variance (ANOVAs) were used to examine systematic performance differences across groups and experiments. The raw data associated with this study can be found at the Open Science Framework (<https://osf.io/u75zh/>).

2.2 | Results and discussion

Both adults and children chose the triplet sequence more often than chance in the 2AFC test (Adults: $t(29) = 6.836, p < 0.001, d = 1.248$; Children: $t(29) = 3.255, p = 0.001, d = 0.594$; Figure 1c). We observed that adults' performance (mean accuracy = 75.42%, SD = 20, 36) was higher than children's (mean accuracy = 64.38%, SD = 24.19), but these differences were not statistically significant, $t(29) = 1.818, p = 0.079, d = 0.332$. Given noticeable difference in adults and children's performance, we also ran a Bayesian test. A test of group difference produced a Bayes Factor of 0.833, indicating that our data provide evidence 1.2 times (inverse of 0.833) more in favor of null hypothesis, stating that there is no difference between the two groups. This result indicates that not only adults (Brady & Oliva, 2008), but also children are capable of extracting item-level regularities from naturalistic stimuli.

These findings show that both adults and children show learning of statistical structure presented with naturalistic stimuli. As we observed that both adults and children can successfully extract *item-level* statistical structure from naturalistic stimuli, we investigated in Experiment 2 whether this learning operates at the *category level* as well.

3 | EXPERIMENT 2

3.1 | Methods

3.1.1 | Participants

Thirty adults (18–23 years, mean age = 19.37, SD = 1.26, 25 female) and 30 children (6–9 years, mean age = 7.9, SD = 1.18, 13 female) participated. Adult participants were recruited from the University of Toronto and compensated with cash or course credit. Children participated at the *Ontario Science Centre* and were compensated with a small toy. All participants (and children's parents) consented

prior to participation, and the experiment was approved by the Research Ethics Boards of the University of Toronto.

3.1.2 | Apparatus, stimuli & procedure

The apparatus, stimuli, and procedure were identical to Experiment 1 except for the following differences: every time an image appeared either during the exposure phase or tests, it was a different and novel image from that animal category. Therefore, the same triplet (e.g., elephant-chicken-bear) appeared each time with different images of elephants, chickens, and bears (see Figure 2a). Thus, triplets operated only at the category level, not at the item level since specific items were only presented once.

Four hundred eight animal images were randomly selected for each participant (288 for exposure, 96 for the 2AFC test, and 24 for the triplet completion test). The triplet structure was constructed by randomly assigning the three animals to each triplet, and this structure was randomly generated for each participant. As in Experiment 1, neither individual triplets nor a pair of triplets appeared in succession during exposure. Notably, all of the images used in the 2AFC test and the triplet completion tests were different from those that were presented during exposure.

3.2 | Results and discussion

As shown in Figure 2, both adults and children performed better than chance on the 2AFC test. Adults showed a mean accuracy of 60.63% (SD = 18.43), which was different from chance, $t(29) = 3.157$, $p = 0.002$, $d = 0.576$, and children showed a mean accuracy of 58.13% (SD = 17.23), which also differed from chance, $t(29) = 2.583$, $p = 0.007$, $d = 0.594$. Adults' and children's performance did not differ, $t(29) = 0.573$, $p = 0.57$, $d = 0.14$. These results indicate that both adults and children can extract category-level regularities.

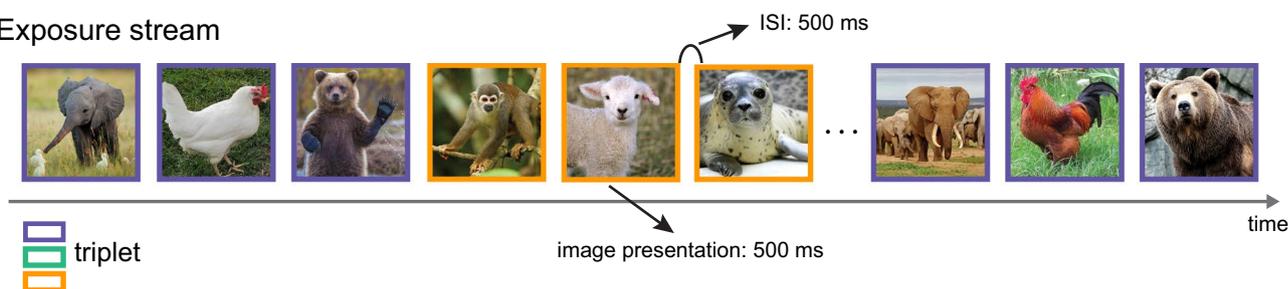
Having established learning at both item (Experiment 1) and category (Experiment 2) levels, we investigated whether learning operates at the category level when participants are exposed to the same items with each presentation of the triplets in Experiment 3.

4 | EXPERIMENT 3

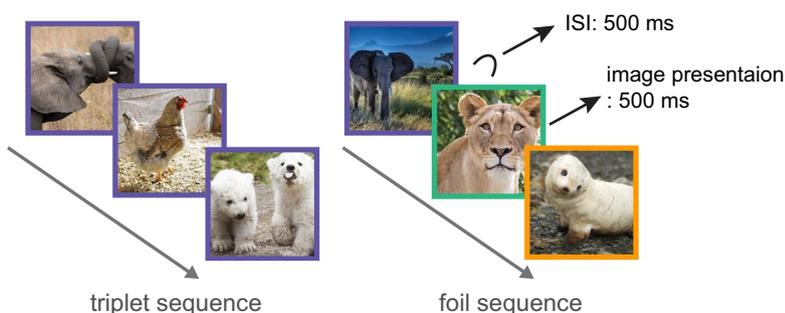
In Experiment 3, the regularities during the exposure phase were constructed with identical images from each category (item-level; same as in Experiment 1), but learning was measured with novel images from the same categories (category-level; same as Experiment 2). If category-level structure is extracted, participants should be

Statistical learning of category-level information (Exp 2)

(a) Exposure stream



(b) 2AFC test



(c) 2AFC performance

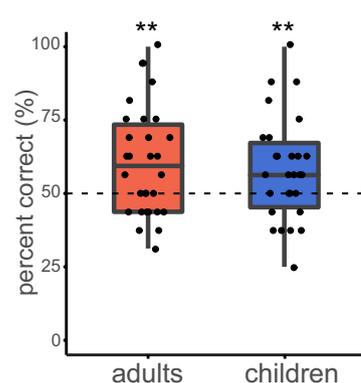


FIGURE 2 Design and results for Experiment 2. (a) Samples from a stream during the exposure phase in Experiment 2. (b) Example of the test stimuli in the 2AFC test phase. Each test image was a novel exemplar from different animal categories. (c) Task performance in the 2AFC test; percentage of triplet chosen as familiar in adults and children

able to generalize their learning from a particular item of each category to different novel items that belong to the same categories.

4.1 | Methods

4.1.1 | Participants

Thirty adults (18–23 years old, mean age = 19.82, SD = 2.1, 25 female) and 30 children (6–9 years old, mean age = 8.01, SD = 1.36, 13 female) participated. Adult participants were recruited from the University of Toronto and compensated with cash or course credit. Children participated at the Ontario Science Centre and were compensated with a small toy. All participants (and children's parents) consented prior to participation, and the experiment was approved by the Research Ethics Boards of the University of Toronto.

4.1.2 | Apparatus, stimuli, and procedure

The exposure phase was identical to Experiment 1: identical images from each category were repeatedly presented in

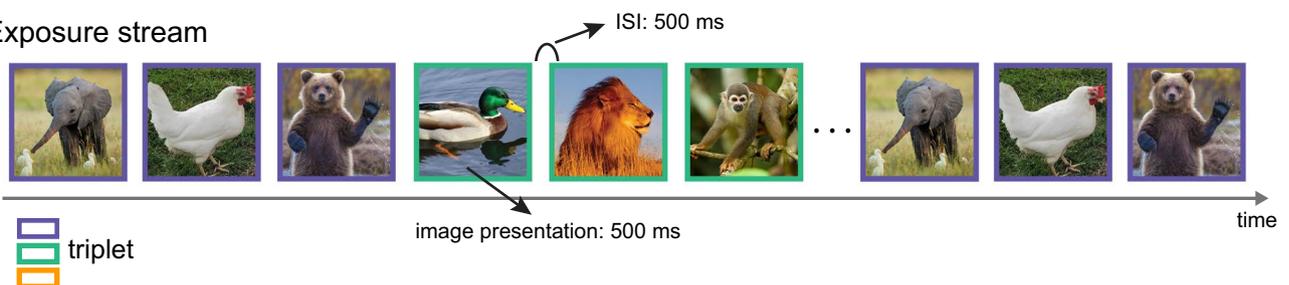
a triplet structure. For example, when there was a triplet of “elephant-chicken-bear,” participants saw exactly the same elephant, the same chicken, and the same bear each time. Thus, the sequence of images encompassed statistical regularities at both, the level of item and the level of category (Figure 3a). However, during the tests, every time an image appeared, it was a novel image from each animal category (same as in Exp2; see Figure 3b).

4.2 | Results & discussions

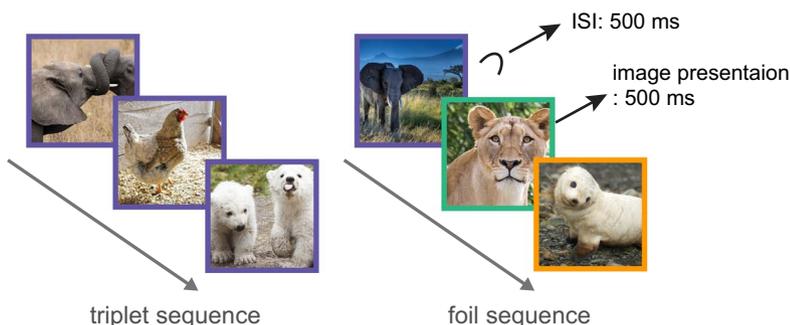
As shown in Figure 3c, both adults and children show evidence of learning on the 2AFC test. Adults showed a mean accuracy of 66.3% (SD = 23.87), which was greater than chance, $t(29) = 3.728$, $p < 0.001$, $d = 0.681$, and children showed a mean accuracy of 59.4% (SD = 16.4), which was also greater than chance, $t(29) = 3.132$, $p = 0.002$, $d = 0.572$. Adults and children did not differ in their learning performance on the 2AFC test, $t(29) = 1.3$, $p = 0.199$, $d = 0.336$. These results indicate that exposure to the same items was sufficient to build category-level structure for *both* adults and children.

Item-learning to category-level generalization (Exp 3)

(a) Exposure stream



(b) 2AFC test



(c) 2AFC performance

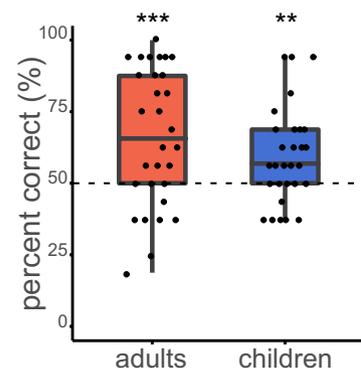


FIGURE 3 Results from Experiment 3. (a) Examples of stimuli stream in the exposure phase. (b) Examples of the test sequences in the 2AFC test. (c) Task performance in the 2AFC test; percentage of triplets chosen as familiar in adults and children



5 | EXPLORATORY ANALYSES: THE TRIPLET COMPLETION TEST

Although adults and children both can learn category-level structure (Exp2 & Exp 3), their ability to express this knowledge might vary depending on retrieval demands given that explicit memory continues to mature over childhood (Ghetti & Angelini, 2008). In other words, with higher retrieval demands, children might not be able to express their learning. To explore this in both children and adults, we performed an overt completion test after the 2AFC test in all experiments.

5.1 | Methods

5.1.1 | Procedure

In this test, participants were presented with the first two items of a triplet and were asked to select the final item to complete it. All four images that appeared at the third position for all of the four triplets were presented as the answer choices at the bottom of the screen. Participants were instructed to click the image that would complete the triplet. Participants could change their response by clicking on a different picture from the choices. The order of the images at the bottom of the screen was randomly shuffled on each trial.

5.1.2 | Data analysis

A Shapiro–Wilk test showed that—across all three experiments—the data from the triplet completion test is not normally distributed, $W = 0.9015$, $p < 0.001$. Thus, we performed Wilcoxon rank order tests and Spearman correlations. To test whether participants showed evidence of learning in the triplet completion test, we performed an one-sample Wilcoxon rank sum test to compare performance to chance level, which was $1 \frac{1}{4}$ (chance in each trial)* 4 trials). The effect size of Wilcoxon test (r) was calculated by determining the z value from the test and dividing by square root of the sample size (Rosenthal, 1994). All Wilcoxon tests were one-tailed, probing whether observed performance was better than chance. To compare performance in adults and children, we performed Spearman rank order correlation.

5.2 | Results

5.2.1 | Experiment 1: item-level learning

Both adults and children completed more triplets than would be expected by chance alone in Experiment 1 (Figure 4a); for adults, mean performance was 2.000, $SD = 1.365$, $W = 246$, $p < 0.001$, $r = 0.642$; and for children, mean performance was 1.933, $SD = 1.258$, $W = 192$, $p < 0.001$, $r = 0.637$. Adults' and children's

performance did not differ, $W = 464$, $p = 0.4189$, $r = 0.147$, mean difference = 0.0667, indicating that both adults and children can express their knowledge about item level statistical structure with higher retrieval demands.

Furthermore, performance on the 2AFC and the triplet completion tests was significantly correlated in both adults (Spearman $\rho = 0.599$, $p < 0.001$) and children (Spearman $\rho = 0.513$, $p = 0.004$) (Figure 4b); thus, both children and adults can equally express their knowledge about item-level statistics regardless of retrieval demands.

5.2.2 | Experiment 2: category-level learning

Adults completed 1.9 triplets ($SD = 1.348$) correctly on average, which is better than chance, $W = 220.5$, $p < 0.001$, $r = 0.665$. However, children correctly completed an average of 1.3 ($SD = 0.952$) triplets, which is not better than chance, $W = 120$, $p = 0.052$, $r = 0.35$, suggesting that children did not show evidence of learning on this test, where retrieval demands were higher. We further observed that adults' performance was significantly better than children's, $W = 562$, $p = 0.043$, $r = 0.368$, mean difference = 0.6, thus suggesting that adults, but not children, can express learning at the category level with higher demands of retrieval.

Task performance on the triplet completion test correlated with the performance on the 2AFC test in adults, $\rho = 0.443$, $p = 0.007$, but not in children, $\rho = -0.022$, $p = 0.547$. Thus, adults are able to express their learning similarly on these two measures, which differ in terms of their retrieval demands, while children are only able to express their learning when demands are reduced.

5.3 | Experiment 3: generalization

In Experiment 3, adults correctly completed 1.933 triplets on average, $SD = 1.258$, which was better than chance, $W = 227$, $p < 0.001$, $r = 0.64$. However, children did not show evidence of learning; they completed an average of 1.267 correctly, which was not better than chance (1.03), $SD = 0.94$, $W = 130$, $p = 0.064$, $r = 0.33$. Adults' performance was better than the children's performance, $W = 585$, $p = 0.019$, $r = 0.425$, mean difference = 0.667, indicating that children could not express their learning at the category level with higher demands of retrieval.

The performance on the 2AFC test and the triplet completion test was correlated in adults, $\rho = 0.449$, $p = 0.013$, but not in children, $\rho = 0.069$, $p = 0.717$. Thus, adults are able to express their learning similarly with these two measures, which differ in terms of their retrieval demands, while children are only able to express their learning with less demands of retrieval, on the 2AFC test.

Together, the triplet completion test demonstrates that adults and children have different abilities to express their learning under varying degree of retrieval demands when their knowledge was tested at the level of category.

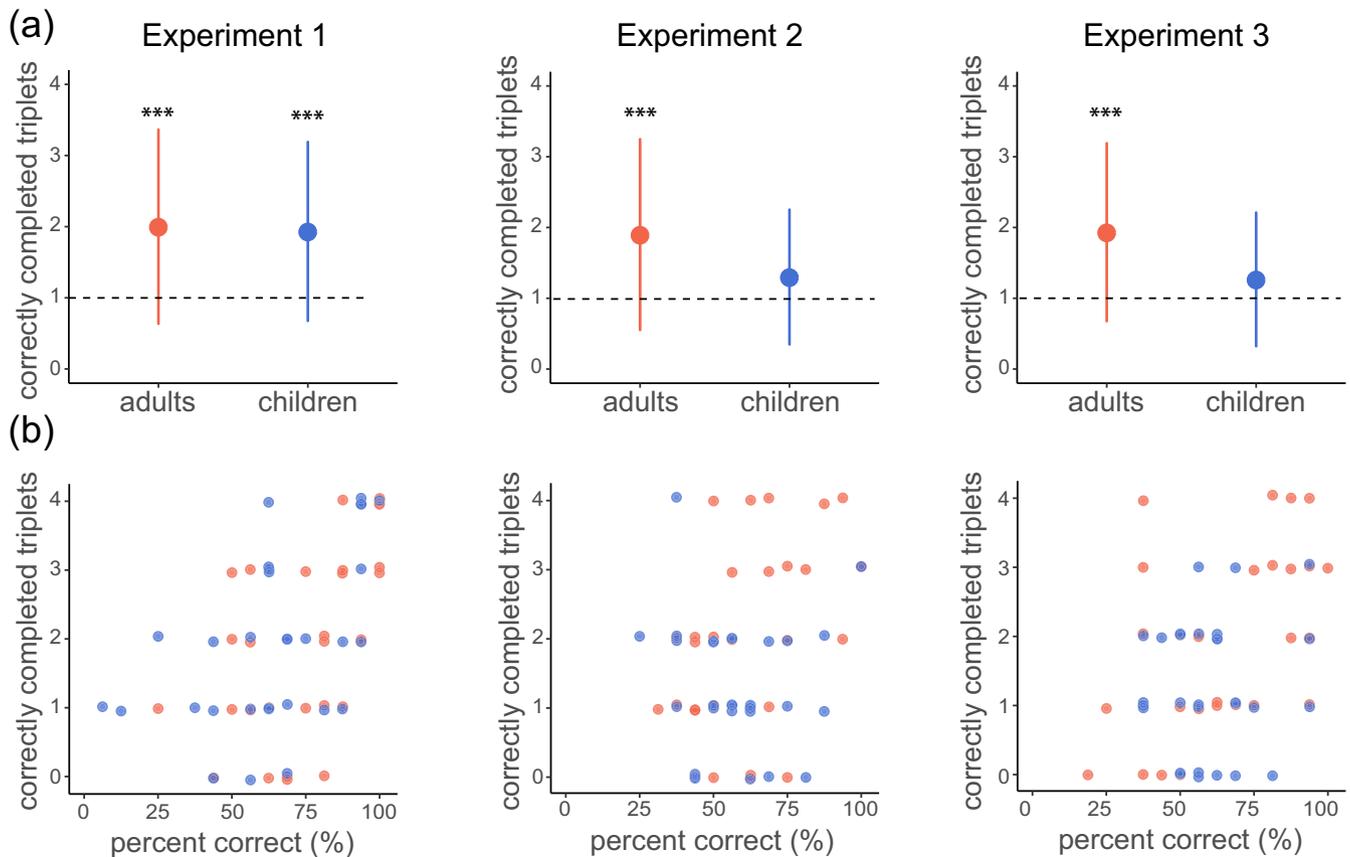


FIGURE 4 (a) Task performance in the triplet completion test in Experiment 1-3. Mean performance is marked with a filled circle, and the bar indicates the $\pm 1SE$ from the mean. Statistical significance is marked (***) $p < 0.001$ for the comparison of the mean performance and the chance level (highlighted in a dotted line). (b) Scatter plots of task performance of the 2AFC test and the triplet completion test in Experiment 1-3. Spearman correlation between the two measures in each group is indicated above the regression line

6 | COMPARING PERFORMANCE ACROSS EXPERIMENTS

6.1 | Confirmatory test (2AFC)

An ANOVA with experimental condition (Exp1, Exp2 and Exp3) and age (child, adult) as variables, revealed a main effect of experiment, $F(2,174) = 4.186$, $p = 0.017$, $\eta^2 = 0.046$, and a main effect of age, $F(1,174) = 5.053$, $p = 0.026$, $\eta^2 = 0.028$ on 2AFC performance, but no interaction between the two factors, $F(2, 174) = 0.663$, $p = 0.516$, $\eta^2 = 0.007$, Figure 5a. In other words, across all experiments, adults performed better than children. Also, both adults and children performed the best in Exp1 where the exposure and the test were conducted with the identical stimuli, as compared to Exp2 and Exp3, where the exposure and the test were comprised of different sets of the stimuli.

6.2 | Exploratory test (Triplet Completion test)

For the triplet completion test, shown in Figure 5b, an ANOVA with experimental condition and age as variables revealed a main effect of age, $F(1,174) = 6.172$, $p = 0.014$, $\eta^2 = 0.0343$, but no main effect of

experiment, $F(2,174) = 1.867$, $p = 0.158$, $\eta^2 = 0.021$, suggesting that adults performed better than children across the three experiments. There was no interaction between age and experimental condition, $F(2,174) = 1.126$, $p = 0.327$, $\eta^2 = 0.013$.

Although there was no interaction between age and experimental condition, adults and children show different learning patterns across experiments. Adults showed robust evidence of learning both when they were tested with the identical items that they were exposed to (Exp1; mean performance = 2) and when they were tested with novel items (as in Exp2, mean performance = 1.9), regardless of whether exposure consisted of unique items (in Exp2) or whether exposure was constructed with the same items with each presentation of a triplet (as in Exp3; mean performance = 1.933). Thus, adults are able to express their learning on this more direct test (with higher retrieval demands) regardless of the level at which information was presented during learning and test.

Children, however, showed robust evidence of learning in triplet completion performance only when they were tested with the identical items that they had previously seen (Exp1). When they were tested with novel items, they did not show evidence of learning, regardless of whether exposure included the structure constructed with multiple novel items (Exp2), or whether exposure was

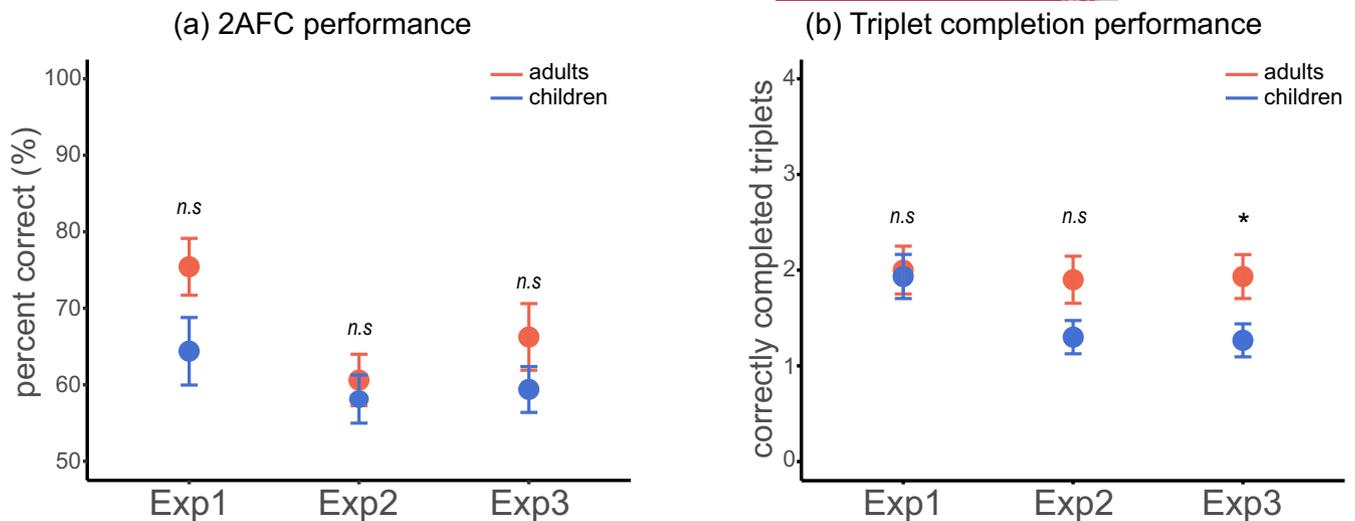


FIGURE 5 (a) 2AFC performance in adults and children across experiments. (b) Triplet completion test performance in adults and children across experiments. Mean performance is marked with the filled circle separately for adults (red) and children (blue), and error bars indicate the ± 1 SE from the mean. Statistical significance is marked ($*p < 0.05$) for the comparison between the age groups

constructed with consistently the same items (Exp3). Thus, on this test where retrieval demands were high, children were only able to express their learning when the same images were repeatedly shown during exposure and during test.

7 | GENERAL DISCUSSION

The current study explores whether item- or category-level regularities are prioritized in children's statistical learning and how children's learning differs from adults'. As in adults, we find that children can learn regularities operating at multiple levels simultaneously. In the first two experiments, we echo previous work showing that statistical learning can occur with naturalistic stimuli at the item level (Exp1) and the category level (Exp2) in adults (Brady & Oliva, 2008). For the first time, we show that young children are capable of learning statistical information at a category level. We further show that statistical learning can spread to novel exemplars that observers had never encountered previously, even when exposure consisted of specific items, in adults and children alike (Exp 3).

The current study replicates findings from the previous studies in adults (Brady & Oliva, 2008; Otsuka et al., 2013), showing that observers can learn at the category level even when the learning experience consists of specific items from the category (that is, the same items that are experienced over and over). Although these previous studies have shown that learning from specific items can generalize to the semantic level using lexical items (Brady & Oliva, 2008; Otsuka et al., 2013), it had not been explicitly tested whether this generalization allows observers to extend their learning to novel visual exemplars that share category-level regularities. The current study extends the previous findings by showing that statistical learning can indeed generalize to novel items that observers have never encountered.

The present study critically extends our understanding of statistical learning in children. For the first time, we show that children's statistical learning can occur at the level of categories, even when they are exposed to a structure that operates over items. These findings fit with previous research, which shows that children can generalize knowledge obtained from observed stimuli to novel stimuli that follow the same temporal contingency rules (Gomez & Gerken, 1999) or those that share the same category membership (Gelman & Coley, 1990). Here we add that children can benefit from their category knowledge in the context of statistical learning.

In possible contrast with our findings, previous work has suggested that children might be preferably oriented to item-level information. Along these lines, Sloutsky and Fisher (2004) have shown that when making inferences about the pictures of animals, children were more oriented to items themselves, rather than their categories, remembering individual items better than adults when asked to think about the objects at the level of the category. Furthermore, children are equally sensitive to features that are diagnostic to category membership and those that are not, whereas adults prioritize diagnostic features when learning categories (Deng & Sloutsky, 2015; Savic & Sloutsky, 2019). In the current study, however, we did not find support for an orientation toward item-level information in children in the domain of statistical learning. Children's 2AFC test performance indicated that they were able to learn both the item- and category-level information.

In contrast, our exploratory measure showed that children are uniquely able to express their learning on the triplet completion test when items are the same during exposure and test (Exp 1), whereas adults can express their learning on this test similarly across the three experiments. Given that the triplet completion task requires the overt identification of a member of a triplet, the retrieval demands for this task are higher than for the 2AFC test. Children appear to be less able to express their learning with higher retrieval

demands specifically when either exposure or test includes *novel* exemplars. These findings support the idea that statistical learning might operate differently in adults and children (Finn et al., 2019; Gómez, 2017; Smith et al., 2018), and this seems to be especially true for levels of information that are categorical (and more abstract). However, these potential differences in adults and children observed in the triplet completion test need to be carefully interpreted for the following reasons. First, it should be noted that there was no significant interaction between group (child or adult) and the experimental conditions on the triplet completion test, suggesting that the differences in adults and children across experiments may be negligible. Furthermore, it should be noted that the triplet completion task was exploratory and has limited sensitivity for measuring learning given that it was comprised of only four trials. Finally, because this test was performed at the end of the experiment, children's performance might have been impacted by fatigue. Thus, further research is needed to better understand how statistical learning may be expressed differently in adults and children when retrieval demands vary.

Our findings demonstrate that adults show better learning performance than children across the three experiments on both the 2AFC and exploratory triplet completion tests. While our observation of superior learning in adults appears to conflict with some previous research, which shows that similarly aged (6- to 7-year-old) children's statistical learning performance did not differ from adults' during an auditory statistical learning task that used syllables (Saffran et al., 1997), it fits with other work showing that adults display superior performance (as compared to children aged 6–11 years and adolescents aged 12–17 years) on a visual statistical learning task (Schlichting et al., 2016). Studies looking at changes within a broad age range during childhood (5–12 years) also tend to show improvement with age (Arciuli & Simpson, 2011; Raviv & Arnon, 2018; Shufaniya & Arnon, 2018). Indeed, the age-related changes in statistical learning appear to be related to the specific stimuli and sensory modality that are used: when linguistic (and auditory) stimuli are used, such as syllables as in the Saffran et al., 1997, no age-related differences tend to be observed (Finn et al., 2019; Shufaniya & Arnon, 2018). The age difference observed in our study therefore is well aligned with the previous research given the visual (and non-linguistic) nature of our stimuli. Unlike the previous studies, however, we did not observe age-related changes within the group of children (Figure S1). This is likely due to our relatively limited age range within childhood (6–9 years). Still, our findings of superior learning in adults replicate previous work on developmental changes in statistical learning and suggest that the mechanisms underlying statistical learning may continue to develop across childhood.

Further work is needed to explore the limits of adults' and children's ability to extend their statistical learning from items to more abstract categories. Several studies have suggested that there are constraints on how far statistical learning can be generalized in adults. For example, learning does not transfer from items (e.g., apple–bus) to superordinate categories (e.g., any fruit–any vehicle) (Otsuka et al., 2014; also see Luo & Zhao, 2018). Furthermore, when

the test consisted of both the exposed items (having both item- and category-level information) and the novel items from the exposed category (having only category-level information), adult observers showed a higher sensitivity to items over categories. These findings imply that item-level information is intact and possibly prioritized when learning category-level regularities in adults (Emberson & Rubinstein, 2016; Jun & Chong, 2018). However, we do not know yet whether there are limits to generalization in children. Indeed, it is possible that children's statistical learning might be more flexibly transferred to the category level given their limited ability to encode and distinguish specific instances (Bauer & Dow, 1994; Brown & Scott, 1971). Of course, the opposite (more limits) could also be true given the previously noted work on children's better memory for and a greater focus on items in certain circumstances (Sloutsky & Fisher, 2004). Further research is required to address this.

Category knowledge may also play a critical role in statistical learning at the category level. In the current study, all of the categories used were familiar to both groups, and thus, verbalizable, which could critically impact category-level statistical learning both in adults and children. That is, verbally labeling each animal from the picture might be what enables category-level learning. To explore this important question about how category-level learning occurs, we asked adult participants to think back on their experience and indicate how often they verbalized each animal during the exposure phase. While there was significant variability in their verbalization frequency (see Figure S2), verbalization frequency during the exposure did not relate to 2AFC performance at the category level in Exp2 and Exp3 (Figure S2). Although our data do not indicate that the frequency of verbalization is related to learning, verbalization could still be critical for statistical learning at the category level. To understand this further, designs that suppress verbalization or use categories without verbal labels are needed. For example, if participants were asked to engaged in a verbal task (e.g., counting) during the exposure phase, it should interfere with category labeling and would thus shed light on the role of labeling in category-level statistical learning.

The data reported here are clear: children, like adults, can build representations of both item- and category-level information during statistical learning, even when they are merely exposed to the same items over and over again. These findings suggest that children's statistical learning is more flexible than previously thought—learning can extend to items that were never seen before. Notably, we have shown that expectations about distributions can be formed for categories of items that a learner has never seen, something that has been noted as core to faculties from understanding a languages' grammar (Slobin, 1973) to predicting future events (Bar, 2007). Furthermore, our findings from Exp1 and Exp3 together demonstrate that children can extract more than one kind of structure at a time, which has so far been shown only in adults (Brady & Oliva, 2008; Emberson & Rubinstein, 2016; Turk-Browne et al., 2008). Given that most of the naturalistic stimuli that we encounter in real life have information present at multiple levels simultaneously, these findings suggest that

statistical learning could shape how children organize conceptual knowledge, contributing to important processes, such as building schemas and forming expectations about future events.

ACKNOWLEDGEMENTS

We would like to thank Aria Wills, Theresa Palm, and Danielle Lim for their help with data collection. This work was supported by the Social Sciences and Humanities Research Council of Canada [Insight Program, 435171493 to ASF; 430-2017-01189 to DBW], Canada Foundation for Innovation and Ontario Research Fund (34947 to ASF), and the Natural Sciences and Engineering Research Council of Canada [RGPIN-2016-05 to ASF; RGPIN-2015-06696 to DBW].

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Open Science Framework at [10.17605/OSF.IO/U75ZH](https://doi.org/10.17605/OSF.IO/U75ZH).

ORCID

Yaelan Jung  <https://orcid.org/0000-0002-9047-2583>

Dirk B. Walther  <https://orcid.org/0000-0001-8585-9858>

REFERENCES

- Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464–473. <https://doi.org/10.1111/j.1467-7687.2009.00937.x>
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. <https://doi.org/10.1016/j.tics.2007.05.005>
- Bauer, P. J., & Dow, G. A. (1994). Episodic memory in 16- and 20-month-old children: Specifics are generalized but not forgotten. *Developmental Psychology*, *30*(3), 403. <https://doi.org/10.1037/0012-1649.30.3.403>
- Bjorklund, D. F. (1988). Acquiring a mnemonic: Age and category knowledge effects. *Journal of Experimental Child Psychology*, *45*(1), 71–87. [https://doi.org/10.1016/0022-0965\(88\)90051-3](https://doi.org/10.1016/0022-0965(88)90051-3)
- Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, *19*(7), 678–685. <https://doi.org/10.1111/j.1467-9280.2008.02142.x>
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Brown, A. L., & Scott, M. S. (1971). Recognition memory for pictures in preschool children. *Journal of Experimental Child Psychology*, *11*(3), 401–412. [https://doi.org/10.1016/0022-0965\(71\)90045-2](https://doi.org/10.1016/0022-0965(71)90045-2)
- Deng, W. S., & Sloutsky, V. M. (2015). The development of categorization: Effects of classification and inference training on category representation. *Developmental Psychology*, *51*(3), 392. <https://doi.org/10.1037/a0038749>
- Emberson, L. L., & Rubinsten, D. Y. (2016). Statistical learning is constrained to less abstract patterns in complex sensory input (but not the least). *Cognition*, *153*, 63–78. <https://doi.org/10.1016/j.cognition.2016.04.010>
- 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. <https://doi.org/10.1016/j.cognition.2008.04.002>
- Finn, A. S., Kharitonova, M., Holtby, N., & Sheridan, M. A. (2019). Prefrontal and hippocampal structure predict statistical learning ability in early childhood. *Journal of Cognitive Neuroscience*, *31*(1), 126–137.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. <https://doi.org/10.1111/1467-9280.00392>
- Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, *134*(4), 521. <https://doi.org/10.1037/0096-3445.134.4.521>
- Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. <https://doi.org/10.1016/j.tics.2014.12.010>
- Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology*, *26*(5), 796. <https://doi.org/10.1037/0012-1649.26.5.796>
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183–209. [https://doi.org/10.1016/0010-0277\(86\)90034-X](https://doi.org/10.1016/0010-0277(86)90034-X)
- Ghetti, S., & Angelini, L. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. *Child Development*, *79*(2), 339–358.
- Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160054.
- Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135. [https://doi.org/10.1016/S0010-0277\(99\)00003-7](https://doi.org/10.1016/S0010-0277(99)00003-7)
- Jun, J., & Chong, S. C. (2016). Visual statistical learning of temporal structures at different hierarchical levels. *Attention, Perception, & Psychophysics*, *78*(5), 1308–1323. <https://doi.org/10.3758/s13414-016-1104-9>
- Jun, J., & Chong, S. C. (2018). Visual statistical learning at basic and subordinate category levels in real-world images. *Attention, Perception, & Psychophysics*, *80*(8), 1946–1961. <https://doi.org/10.3758/s13414-018-1566-z>
- Keresztes, A., Bender, A. R., Bodammer, N. C., Lindenberger, U., Shing, Y. L., & Werkle-Bergner, M. (2017). Hippocampal maturity promotes memory distinctiveness in childhood and adolescence. *Proceedings of the National Academy of Sciences*, *114*(34), 9212–9217. <https://doi.org/10.1073/pnas.1710654114>
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. [https://doi.org/10.1016/S0010-0277\(02\)00004-5](https://doi.org/10.1016/S0010-0277(02)00004-5)
- Luo, Y., & Zhao, J. (2018). Statistical learning creates novel object associations via transitive relations. *Psychological Science*, *29*(8), 1207–1220. <https://doi.org/10.1177/0956797618762400>
- Ngo, C. T., Newcombe, N. S., & Olson, I. R. (2018). The ontogeny of relational memory and pattern separation. *Developmental Science*, *21*(2), e12556. <https://doi.org/10.1111/desc.12556>
- Otsuka, S., Nishiyama, M., & Kawaguchi, J. (2014). Constraint on the semantic flexibility in visual statistical learning. *Visual Cognition*, *22*(7), 865–880. <https://doi.org/10.1080/13506285.2014.923548>
- Otsuka, S., Nishiyama, M., Nakahara, F., & Kawaguchi, J. (2013). Visual statistical learning based on the perceptual and semantic information of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 196. <https://doi.org/10.1037/a0028645>

- Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science*, 21(4), e12593. <https://doi.org/10.1111/desc.12593>
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). Russell Sage Foundation.
- Saffran, J. R., Aslin, R. N., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- 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. <https://doi.org/10.1111/j.1467-9280.1997.tb00690.x>
- Savic, O., & Sloutsky, V. M. (2019). Assimilation of exceptions? Examining representations of regular and exceptional category members across development. *Journal of Experimental Psychology: General*, 148(6), 1071. <https://doi.org/10.1037/xge0000611>
- Schlichting, M. L., Guarino, K. F., Schapiro, A. C., Turk-Browne, N. B., & Preston, A. R. (2016). Hippocampal structure predicts statistical learning and associative inference abilities during development. *Journal of Cognitive Neuroscience*, 29(1), 37–51. https://doi.org/10.1162/jocn_a_01028
- Shufaniya, A., & Arnon, I. (2018). Statistical learning is not age-invariant during childhood: Performance improves with age across modality. *Cognitive Science*, 42(8), 3100–3115. <https://doi.org/10.1111/cogs.12692>
- Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development*. Holt, Rinehart & Winson.
- Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, 7, 246–251. [https://doi.org/10.1016/S1364-6613\(03\)00109-8](https://doi.org/10.1016/S1364-6613(03)00109-8)
- Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, 133, 166–188. <https://doi.org/10.1037/0096-3445.133.2.166>
- Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science*, 18, 179–185. <https://doi.org/10.1111/j.1467-9280.2007.01869.x>
- Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, 22(4), 325–336. <https://doi.org/10.1016/j.tics.2018.02.004>
- Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 399–407.
- Younger, B. (1990). Infants' detection of correlations among feature categories. *Child Development*, 61(3), 614–620. <https://doi.org/10.2307/1130948>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Jung Y, Walther DB, Finn AS.

Children automatically abstract categorical regularities

during statistical learning. *Dev Sci*. 2020;00:e13072. [https://](https://doi.org/10.1111/desc.13072)

doi.org/10.1111/desc.13072