Cognition 126 (2013) 156–164

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Redundancy matters: Flexible learning of multiple contingencies in infants

Vladimir M. Sloutsky *, Christopher W. Robinson

Department of Psychology and Center for Cognitive Science, The Ohio State University, United States

ARTICLE INFO

Article history:
Received 28 February 2011
Revised 6 September 2012
Accepted 8 September 2012
Available online 9 November 2012

Keywords:
Infancy
Learning
Attention
Categorization

ABSTRACT

Many objects and events can be categorized in different ways, and learning multiple categories in parallel often requires flexibly attending to different stimulus dimensions in different contexts. Although infants and young children often exhibit poor attentional control, several theoretical proposals argue that such flexibility can be achieved without selective attention. If this is the case, then even young infants should be able to learn multiple dimension-context contingencies in parallel. This possibility was tested in four experiments with 14- and 22-month-olds. Learning of contingencies succeeded as long as there were multiple correlations between the context and the to-be-learned dimension. These findings suggest that infants can learn multiple dimension-context contingencies in parallel, but only when there is sufficient redundancy in the input.

1. Introduction

The ability to learn categories is ubiquitous in human cognition: people can treat appreciably different entities (e.g., different dogs, different tokens of the same word, different situations, or different problems) as variants of the same thing. This is a critically important property of human intelligence: it promotes cognitive economy (as different entities are represented the same way) and application of learned knowledge to novel situations (learning that some cats are carnivores would lead one to believe that other cats are carnivores as well). Furthermore, the ability to categorize appears early in development, with 3-month-olds ably learning perceptual categories (Quinn, Eimas, & Rosenkrantz, 1993). How is this ability achieved?

Starting with the pioneering work of Shepard, Hovland, and Jenkins (1961) and Bruner, Goodnow, and Austin (1956), there has been a wealth of theoretical proposals attempting to answer this question. One idea dating back to Shepard et al. (1961) proved to be particularly influential: category learning requires selective attention – the ability to focus on category- relevant dimensions, while ignoring category-irrelevant dimensions. Many influential models of categorization and category learning have adopted this idea (e.g., Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986), which helped these models explain a wide range of empirical data. However, there is an interesting cost of selectivity – the phenomenon known as learned inattention (e.g., Kruschke & Blair, 2000). Specifically, focusing on category-relevant dimensions results in decreased attention to category-irrelevant dimensions: learning of category K, for which dimension *X* (e.g., shape) is irrelevant makes it more difficult to subsequently learn category M, for which dimension *X* is relevant. However, the idea of selective attention and subsequent learned inattention could be difficult to reconcile with some remarkable aspects of category learning: people learn categories both ably and *flexibly*, learning multiple categories in parallel. For example, people may learn food categories that are likely to be organized by color and texture and toy categories that are likely to be organized by shape. Furthermore, dimensions that are relevant for the former categorization could be irrelevant for the latter categorization.

Even more puzzling, young children also learn multiple categories in parallel, and depending on a situation, attend

---

* Corresponding author. Address: Center for Cognitive Science, Cognitive Development Lab, 1961 Tuttle Park Place, The Ohio State University, Columbus, OH 43210, United States. Tel.: +1 614 688 5855; fax: +1 614 292 0321.

E-mail addresses: sloutsky.1@osu.edu (V.M. Sloutsky), robinson.777@osu.edu (C.W. Robinson).

0010-0277/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2012.09.016

to different dimensions (e.g., Macario, 1991; Opfer & Bulloch, 2007; Sloutsky & Fisher, 2008). For example, even preschoolers may attend to color when categorizing food and attend to shape when categorizing toys (Macario, 1991). Furthermore, infants can categorize the same set of stimuli in different ways (Ellis & Oakes, 2006; Mareschal & Tan, 2007). This flexibility is particularly surprising given that on many other tasks, infants and young children often exhibit inflexibility (see Hanania & Smith, 2010, for a recent review). For example, in the A-not-B task infants often fail to exhibit flexibility by failing to search in a new location after multiple observations of a toy being hidden in a particular location (Diamond, 1985). Similarly, 3-year-olds often perseverate and fail to switch their attention to a new dimension on a Dimensional Change Card Sort task (Zelazo, Frye, & Rapus, 1996).

How is it possible to reconcile the idea of selectivity (and associated costs of learned inattention) with flexible learning of multiple categories? And how is it possible to reconcile children's inflexibility on some tasks with flexibility on other tasks? One idea that may reconcile these findings is that category learning is sub-served by multiple systems (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; see also Love & Gureckis, 2007; Sloutsky, 2010, for reviews). One putative system is evolutionarily primitive and is based on projections from modality-specific cortices to the striatum. This "compression-based" system exploits redundancy in the input without deploying selective attention. This system may underlie implicit and unsupervised learning of highly correlated structures. Because compression-based learning is not based on selectivity, it may bear no costs of selective attention, thus allowing flexible (albeit implicit) learning. Furthermore, because this system exhibits an early developmental onset, this flexibility should be observed early in development.

A second (more evolutionarily recent) system is based on selective attention and involves the prefrontal cortex. The "selection-based" system enables focusing on some dimensions, while ignoring others, which often results in learned inattention to the ignored dimensions. This system may underlie learning of structures that are based on few relevant dimensions. Given that these dimensions are often difficult to discover, such selection-based learning often requires supervision. In contrast to compression-based learning, selective attention is central for this system, which may result in costs of selective attention – learned inattention to formerly ignored dimensions.

The multiple systems framework makes three interesting predictions. First, given that the compression-based system exhibits early onset (see Sloutsky, 2010, for a review), learning that is sub-served by this system should exhibit early onset as well. As a result, even infants who presumably have a functioning compression-based system of learning may be able to exhibit flexible learning on tasks that engage this system. Second, increasing the number of correlated features should facilitate learning in the compression-based system. While the ability of infants to learn correlations is fairly non-controversial and well documented (e.g., Rakison & Poulin-Dubois, 2002; Younger & Cohen, 1983), and there is also evidence that redundant cues can facilitate learning (Thiessen & Saffran, 2003), the novelty of the current study is that it focuses on redundant cues facilitating <u>flexible</u> learning of <u>multiple</u> contingencies. And, third, assuming that infants rely on compression-based learning, there should be no costs that are typically associated with selective attention, such as learned inattention.

Testing these hypotheses was the primary goal of the current research. The current study used a modified contingency learning procedure. Infants in the reported experiments were familiarized to pairs of images (see Fig. 1A). When images appeared in Context 1, they matched on one dimension (e.g., shape), and when images appeared in Context 2, they matched on a different dimension (e.g., color). After training, we used a modified switch task (cf. Younger & Cohen, 1983) to determine if infants learned the arbitrary dimension-context contingencies (see Fig. 1B). On "Same" test trials, the dimension-context relation was identical to training. On "Switch" trials, the familiar dimensions and familiar contexts switched (e.g., the dimension familiarized in Context 1 was now presented in Context 2).

We hypothesized that when the to-be-learned dimension correlates with multiple context features, infants would learn both contingencies in parallel, with little or no cost of selectivity (Experiment 1). Such learning would require flexibility because infants need to switch between learning of shape in one context and color in a different context. Furthermore, assuming that the current task engages the compression-based system (because no explicit instructions or feedback were provided), learning of dimension-context contingencies should decrease in contexts that provide fewer correlated features (Experiments 2–3).

2. Experiment 1

The goal of Experiment 1 was to test the hypothesis that infants could learn two different dimension-context contingencies in parallel, such that they would flexibly switch between attending to shape in one context and to color in another context.

3. Method

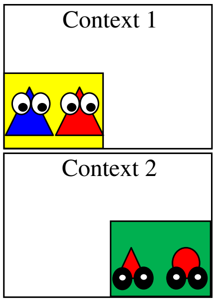
3.1. Participants

Participants were 34 14-month-olds (17 boys and 17 girls, $M = 459.47$ days, $SE = 10.38$ days) and 34 22-month-olds (17 boys and 17 girls, $M = 672.15$ days, $SE = 10.85$ days). In this and other experiments reported here, parents' names were collected from local birth announcements, and contact information was obtained through local directories. All children were full-term (i.e., >2500 g birth weight) with no auditory or visual deficits, as reported by parents. A majority of infants were Caucasian. An additional seven infants were tested but not included in the final sample due to fussiness.

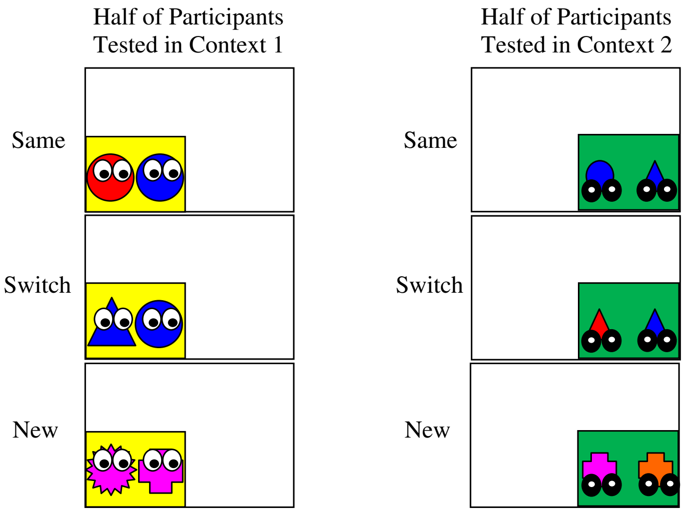
3.2. Apparatus

Infants were seated on parents' laps approximately 100 cm away from a 152 cm × 127 cm projection screen.

(A) Training Phase for All Participants in Experiment 1



(B) Testing Phase in Experiment 1: Same, Switch, and New Blocks



(C) Time Course Within a Training or Testing Block

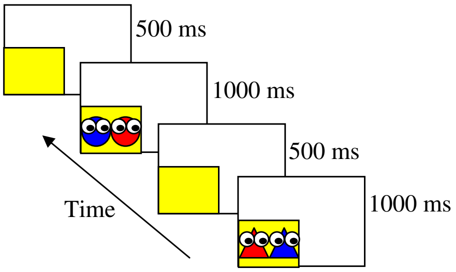


Fig. 1. (A) Examples of training blocks. (B) Examples of testing blocks. (C) Time course within a training or testing block.

A NEC GT2150 LCD projector was mounted to the ceiling approximately 30 cm behind the infant. A Dell Dimension 8200 computer with *Presentation* software was used to present stimuli and record visual fixations. Fixations were recorded online by pressing a USB gamepad when infants were looking at the stimulus and by releasing the button when infants looked away from the stimulus. Two video streams (i.e., stream of stimulus presentation and stream of infants' fixations) were projected onto two Dell flat panel monitors in an adjacent room, and a Sony DCR-PC120 camcorder recorded both video streams. This split-screen recording was used to establish inter-rater reliability. Thirty percent of the infants were coded offline and reliability between online and offline coders across all reported experiments was acceptable, $r = .96$.

3.3. Stimuli

Training and testing items varied in shape (circle or triangle) and color (red or blue). The image pairs matched in shape in one context and matched in color in a different context, and the dimension-context pairing was counterbalanced across participants. Assigning two shapes and two colors to the images resulted in eight different pairings

Table 1
Abstract structure of stimulus pairs. "Sh" = shape and "C" = color. Values 0 and 1 denote circle and triangle for the Shape dimension and red and blue for the Color dimension.

	Category 1 (shape match)				Category 2 (color match)			
	Stimulus 1		Stimulus 2		Stimulus 1		Stimulus 2	
Stimulus pair	Sh	C	Sh	C	Sh	C	Sh	C
Stimulus pair 1	1	1	1	0	1	1	0	1
Stimulus pair 2	0	0	0	1	0	0	1	0
Stimulus pair 3	1	1	1	0	0	1	1	1
Stimulus pair 4	0	0	0	1	1	0	0	0

across the two contexts. As can be seen in Table 1 the categories were mutually exclusive, thus, during training the same infant would never see items matching in shape in Context 1 and in Context 2. Also, as can be seen in the table, each feature and each dimension was equally frequent across the two contexts; therefore, infants could not learn the dimension-context pairings by picking up on differences in variability or frequency of the individual features or dimensions. Rather, to successfully learn both sets of correlated features, infants had to notice that the two images matched on Dimension 1 in Context 1 and matched on Dimension 2 in Context 2 (we revisit this issue in the General Discussion).

The two contexts differed on four dimensions: stimulus size, spatial location of the stimulus items, the color of the background, and the color and spatial location of an extra feature added to the objects (see Fig. 1). Stimuli presented in Context 1 were large (20 × 20 cm, subtended 11.42° of visual angle), presented on the left side of the screen, presented on a 50 cm × 30 cm yellow background, and they had eye-like circles measuring 2 cm × 2 cm attached to the upper part of the stimuli. Stimuli presented in Context 2 were small (10 × 10 cm, subtended 5.73° of visual angle), presented on the right side of the screen, presented on a 50 cm × 30 cm green background, and they had wheel-like circles measuring 2 cm × 2 cm attached to the lower part of the stimuli.

3.4. Procedure

The procedure consisted of two phases: Training and Testing. There were four training blocks: In two of the blocks images pairs were presented in Context 1, and in the other two blocks, image pairs were presented in Context 2 (see Fig. 1A for examples of matching shape and matching color blocks). The onset of each block started when the infant fixated on a fixation stimulus, and within each block, stimulus pairs were presented multiple times for 1000 ms each, followed by a 500 ms inter-stimulus interval. To ensure that infants had an opportunity to see images in one context prior to shifting between contexts, the first two blocks were 36 s in duration and the last two blocks were 24 s in duration. The orders of the first two training blocks and the last two training blocks were randomized for each infant.

The Training phase was followed immediately by the Testing phase, in which half of the participants were tested on shape and half on color. There were three testing blocks (each lasting 24 s in duration): Same, Switch, and New. In the Same block, the dimension-context relation was identical to training. In the Switch block, the familiar dimensions and familiar contexts switched (e.g., the dimension familiarized in Context 1 was now presented in Context 2). In the New Block, there were completely novel images and a switched dimension-context relation. Testing blocks were pseudo-randomized, such that the order of Same and Switch blocks were randomized for each infant and the New block was always presented last.

While encoding of individual colors and shapes was sufficient for discriminating Same from New, such encoding was not sufficient for discriminating Same from Switch (e.g., matching color and matching shape were equally familiar to young infants). Thus, infants could discriminate Same and Switch blocks by learning the dimension-context contingencies (i.e., items match in shape in one context and match in color in a different context).

4. Results and discussion

Primary analyses focused on accumulated looking to Same and Switch blocks for those infants who increased looking on New trials. However, because 28% of the infants across the reported experiments did not reach this criterion, we separately analyzed the data from this group of infants. Novelty responders across all reported experiments are presented in Table 2 and familiarity responders are presented in Table 3. As can be seen in Table 3, in those conditions that had sufficient number of participants and

Table 2
Only novelty responders are reported in Table 2. Accumulated looking during training and test across the different conditions. Note: "*" denotes that learning was successful – looking in the Switch block was significantly longer than looking in the Same block, p's < .05 and η_p^2 denotes the effect size of learning (as measured by the magnitude of difference in looking in the Switch and Same blocks).

Number of correlated contextual features	Exp.	Dimension tested	Looking during training (SE)	Looking to same (SE)	Looking to switch (SE)	η_p^2
4 Correlated features*	1	Shape	88.38 (3.84)	13.80 (0.96)	17.78 (1.01)	.48
4 Correlated features*	1	Color	86.54 (4.09)	12.95 (1.15)	16.39 (1.41)	.27
4 Correlated features*	2B	Shape	88.97 (2.99)	14.01 (1.45)	18.57 (1.12)	.51
4 Correlated features*	2B	Color	93.39 (3.21)	12.57 (1.39)	15.99 (1.44)	.25
3 correlated features*	2A	Shape	79.28 (3.96)	11.28 (0.90)	13.78 (1.01)	.17
3 Correlated features	2A	Color	79.78 (3.21)	11.34 (1.12)	12.53 (1.07)	.08
1 Correlated feature	3	Shape	81.40 (4.31)	14.11 (0.97)	13.49 (0.95)	.02

Table 3

Only familiarity responders are reported in Table 3. Accumulated looking during training and test across the different conditions. Due to small samples sizes, data are averaged across the four correlated feature conditions (Experiment 1 and Experiment 2B) and averaged across tested dimension (Tested on Shape vs. Tested on Color). Note: “*” denotes that learning was successful – looking in the Same block was significantly longer than looking in the Switch Block, $p < .05$ and η_p^2 denotes the effect size of learning (as measured by the magnitude of difference in looking in the Switch and Same blocks).

Number of correlated contextual features	Exp.	Dimension tested	Looking during training (SE)	Looking to same (SE)	Looking to switch (SE)	η_p^2
4 Correlated features*	1 & 2B	Shape or Color	84.74 (2.95)	18.40 (0.79)	16.17 (0.77)	.18 ($N = 38$)
3 Correlated features*	2A	Shape or Color	80.19 (5.41)	15.23 (1.20)	12.17 (1.09)	.25 ($N = 21$)
1 Correlated feature	3	Shape	91.37 (5.42)	18.61 (1.10)	15.96 (1.81)	.26 ($N = 9$)

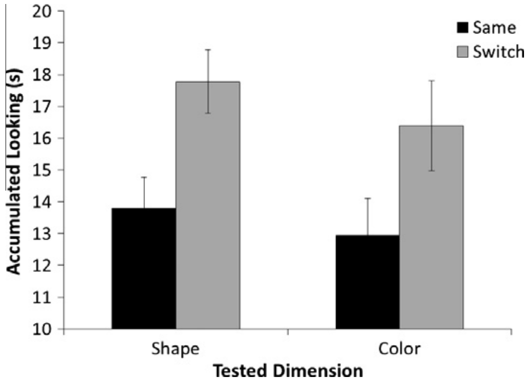


Fig. 2. Mean accumulated looking at test in Experiment 1 as a function of tested dimension and test block. Error bars denote standard errors.

allowed statistical analyses (i.e., the two multiple correlations conditions), familiarity responders were similar to novelty responders (with somewhat weaker learning). However, due to a small sample size in some of the conditions of the familiarity responder group, analyses below focused exclusively on those infants who increased looking on New trials (see Table 2 and Fig. 2).

Mean looking times at test were submitted to an Age (14- vs. 22-months) × Dimension (Tested on Shape vs. Tested on Color) × Test Block type (Same vs. Switch) ANOVA, with Test Block as a repeated measure. The analyses revealed only an effect of Test Block, $F(1, 41) = 22.75$, $p < .001$, $\eta_p^2 = .36$. Specifically, participants accumulated more looking in the Switch block ($M = 17.10$, $SE = 0.86$) than in the Same block ($M = 13.39$, $SE = 0.74$). Furthermore, one-way repeated measures ANOVAs indicated that discrimination of Same and Switch was evident for those infants tested on shape, $F(1, 22) = 20.10$, $p < .001$, $\eta_p^2 = .48$, as well as for those infants tested on color, $F(1, 22) = 7.79$, $p < .01$, $\eta_p^2 = .27$.

To examine potential costs of selectivity, we analyzed effects of order on learning. If infants selectively attend to the dimensions, then they should increase attention to the dimension they encountered first during training and have difficulty switching attention to a new dimension, thus resulting in learning of the first dimension-context contingency and poor learning of the second dimension-context contingency. However, if learning occurs without selective attention, then infants should learn both dimension-context contingencies. The findings are consistent with the latter account. When infants were tested on the dimension-context contingency that was presented during the first 36 s of familiarization, they accumulated more looking in the Switch block ($M = 17.64$, $SE = 1.02$) than in the Same block ($M = 13.84$, $SE = 0.87$), $F(1, 22) = 13.25$, $p < .001$, $\eta_p^2 = .38$. More importantly, infants who were tested on the dimension-context contingency that was presented second during training also accumulated more looking in the Switch block ($M = 16.54$, $SE = 1.40$) than in the Same block ($M = 12.91$, $SE = 1.23$), $F(1, 21) = 11.02$, $p < .005$, $\eta_p^2 = .34$.

These findings support the hypothesis that infants can learn two sets of contingencies in parallel, while exhibiting little evidence of the cost of selectivity. Does a reduction in the number of context features correlating with the target dimension result in an attenuation of learning? This issue was addressed in Experiments 2–3.

5. Experiment 2A

If learning in Experiment 1 was driven by redundancy, then decreasing redundancy should result in attenuated learning. Because the exact number of context features required to support learning is not known, we begin testing this hypothesis by removing one of the correlated context features.

6. Method

6.1. Participants

Participants were 40 14-month-olds (20 boys and 20 girls, $M = 459.13$ days, $SE = 9.17$ days) and 40 22-month-olds (20 boys and 20 girls, $M = 695.18$ days, $SE = 10.49$ days). An additional group of nine infants were tested but not included in the final sample due to fussiness.

6.2. Stimuli and procedure

Experiment 2A was similar to Experiment 1, except that we eliminated one of the contextual cues, while retaining the other three (i.e., image size, spatial location, and color, see Fig. 3B). Stimuli presented in Context 1 were large, presented on the left side of the screen and were presented on a yellow background. Stimuli presented in Context 2 were small, presented on the right side of the screen, and were presented on green background.

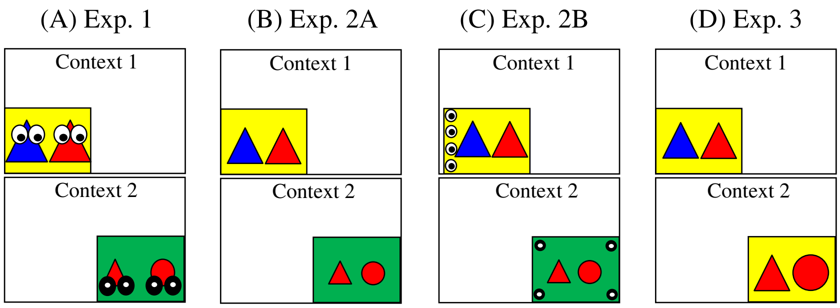


Fig. 3. Example of stimuli across the experiments. (A) Experiment 1. (B) Experiment 2A. (C) Experiment 2B. (D) Experiment 3.

7. Results and discussion

Results from novelty responders are presented in Table 2 and Fig. 4 (see Table 3 for familiarity responders). Mean accumulated looking times at test for novelty responders were submitted to an Age (14- vs. 22-months) × Dimension (Tested on Shape vs. Tested on Color) × Test Block type (Same vs. Switch) ANOVA, with Test Block as a repeated measure. The analyses revealed only an effect of Test Block, $F(1, 55) = 8.01$, $p < .05$, $\eta_p^2 = .13$, with infants accumulating more looking in the Switch block ($M = 13.16$, $SE = 0.73$) than in the Same block ($M = 11.31$, $SE = 0.71$). However, as can be seen in Fig. 4 and in Table 2, this effect was primarily driven by those infants tested on shape: these infants significantly increased looking to Switch test items, $F(1, 29) = 5.95$, $p < .05$, $\eta_p^2 = .17$; whereas, infants tested on Color did not increase looking to Switch test items, $F(1, 28) = 2.45$, $p = .13$, $\eta_p^2 = .08$. Furthermore, greater learning of shape did not stem from infants tested on shape accumulating more looking during training than infants tested on color (see Table 2). These findings suggest that learning decreased compared to Experiment 1: participants learned only one dimension-context contingency, which happened to be based on shape and not color (cf. Casey, 1979).

In summary, when shape and color were correlated with multiple contextual features, infants learned both dimension-context contingencies (Experiment 1), whereas, when there were fewer correlations, participants learned only the shape-context contingency (Experiment 2A). However, it is possible that this difference stems from using potentially meaningful features (e.g., eyes and wheels) in Experiment 1 and removing these features in Experiment 2A. Furthermore, unlike the other contextual features, the eye-like and wheel-like features were part of the target stimuli. We addressed these issues in Experiment 2B by making the features less meaningful and by moving the features to the context.

8. Experiment 2B

If flexible learning in Experiment 1 was driven by infants attending to the meaningful features that were part of the stimulus, then learning should decrease in the current experiment. However, if flexible learning was driven by the number of correlated features, then infants in the current experiment should also flexibly learn the dimension-context contingences because there were comparable numbers of correlated contextual features in Experiments 1 and 2B.

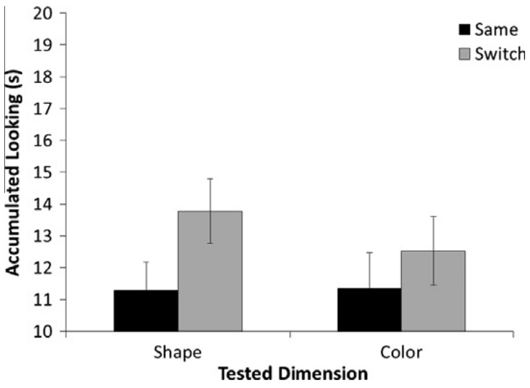


Fig. 4. Mean accumulated looking at test in Experiment 2A as a function of tested dimension and Test Block. Error bars denote standard errors.

9. Method

9.1. Participants

Participants were 28 14-month-olds (14 boys and 14 girls, $M = 451.39$ days, $SE = 12.79$ days) and 28 22-month-olds (14 boys and 14 girls, $M = 670.75$ days, $SE = 12.97$ days). An additional group of six infants were tested but not included in the final sample due to fussiness.

9.2. Stimuli and procedure

Experiment 2B was similar to Experiment 1, except that we moved the eye-like and wheel-like features from the stimuli and placed them on the yellow and green backgrounds (see Fig. 3C). All other stimulus and context features were identical to previous experiments. The procedure was identical to Experiments 1 and 2A.

10. Results and discussion

Results from novelty responders are presented in Table 2 and Fig. 5 (see Table 3 for familiarity responders). An Age (14- vs. 22-months) × Dimension (Tested on Shape

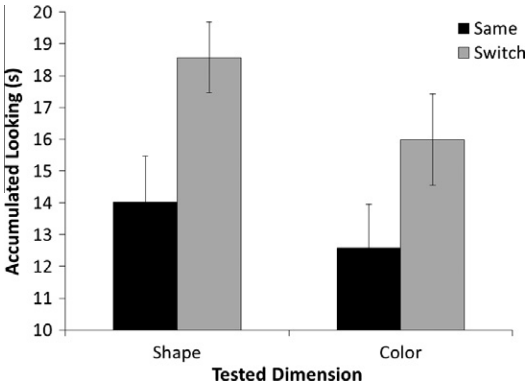


Fig. 5. Mean accumulated looking at test in Experiment 2B as a function of tested dimension and test block. Error bars denote standard errors.

vs. Tested on Color) × Test Block type (Same vs. Switch) ANOVA revealed only an effect of Test Block, $F\ (1, 39) = 23.51$, $p < .001$, $\eta_p^2 = .38$, with infants learning both shape, $F\ (1, 22) = 23.01$, $p < .001$, $\eta_p^2 = .51$, and color, $F\ (1, 17) = 5.53$, $p < .05$, $\eta_p^2 = .25$.

Furthermore, there was no evidence of the cost of selective attention in learning of the dimension-context contingencies: When infants were tested on the dimension-context contingency that was presented during the first 36 s of familiarization, they accumulated more looking in the Switch block ($M = 16.20$, $SE = 1.12$) than in the Same block ($M = 12.80$, $SE = 1.27$), $F\ (1, 20) = 10.37$, $p < .005$, $\eta_p^2 = .34$. More importantly, as in Experiment 1, infants tested on the dimension-context contingency that was presented second during training also accumulated more looking in the Switch block ($M = 17.29$, $SE = 1.37$) than in the Same block ($M = 13.73$, $SE = 1.28$), $F\ (1, 19) = 14.92$, $p < .001$, $\eta_p^2 = .44$. These findings closely replicate Experiment 1 and suggest that infants can flexibly learn multiple dimension-context contingencies; however, as evidenced by Experiment 2A, only when there is sufficient structure in the input.

11. Experiment 3

The goal of Experiment 3 was to further examine the role of redundancy in contingency learning. To address this issue, we left only one context feature correlating with the target dimension. If redundancy is necessary, then learning should further decrease compared to that in Experiment 2A. In addition, given that only the shape-context contingency (but not the color- context contingency) was learned with 3 correlated features in Experiment 2A, the current experiment trained participants on both contingencies, but tested only learning of the shape-context contingency.

12. Method

12.1. Participants

Participants were 20 14-month-olds (10 boys and 10 girls, $M = 486.90$ days, $SE = 8.51$ days) and 20 22-month-olds (10 boys and 10 girls, $M = 674.15$ days, $SE = 13.46$ days). Recall that there was no evidence of infants learning the color-context contingency in Experiment 2A. Therefore, all infants in the current experiment were trained on shape in one context and color in a different context, however, in contrast to previous experiments, infants were only tested on shape in the current experiment. An additional group of four infants were tested but not included in the final sample due to fussiness.

12.2. Stimuli and procedure

Only spatial location distinguished Context 1 from Context 2 (see Fig. 3D). Thus, in Experiment 3, the stimulus pairs presented in Context 1 and Context 2 were identical in size and presented on a yellow background. The rest of the procedure was identical to the shape condition in the previous experiments.

13. Results and discussion

Results from novelty responders are presented in Table 2 and familiarity responders are presented in Table 3. Mean accumulated looking times at test were submitted to an Age (14- vs. 22-months) × Test Block type (Same vs. Switch) ANOVA, with Test Block as a repeated measure. The analysis revealed no significant effects or interactions. Thus, in contrast to previous experiments, the main effect of Test Block did not reach significance, $F\ (1, 30) = 0.48$, $p > .49$, $\eta_p^2 = .02$, with participants showing no evidence of learning. Furthermore, this failure to learn did not stem from infants accumulating less looking to the stimuli during training: accumulated looking during training did not differ between Experiments 2A and 3. These findings suggest that multiple correlated contextual features are necessary for the learning of dimension-context contingencies. However, as can be seen in Table 2, there was a trend with infants in the four-correlated feature conditions accumulating more looking during familiarization than infants in the reduced correlated feature conditions. Thus, it is possible that infants may simply be better learners when presented with a salient context, simply because they were more attentive. According to this account, accumulated looking times during familiarization (measure of stimulus saliency) should be correlated with infants' increased (or decreased, in case of familiarity responders) looking in the Switch block relative to Same block. Across all of the reported experiments, there was no evidence that accumulated looking during familiarization predicted learning: there was no significant correlation between accumulated looking during familiarization and absolute difference in accumulated looking in the Same and Switch blocks, $r\ (243) = .02$, $p > .74$. Therefore, there was little reason to believe that variability in accumulated looking during familiarization was predictive of learning the dimension-context contingencies.

Overall, Experiments 1–3 point to three novel findings. First, infants can learn dimension-context contingencies. Second, they can learn multiple contingencies in parallel. And third, learning depends critically on multiple redundant context features. In particular, effect sizes across the different experiments (see Table 2) suggest that learning was particularly robust when there were multiple correlated contextual features, and it became progressively weaker with the reduction of the number of correlated features.

14. General discussion

Results of the three reported experiments indicate that participants exhibited robust and flexible learning of contingencies when there were multiple correlations between the context and the target dimension (i.e., either color or shape). When the number of correlations between the context and the to-be-learned dimension was somewhat reduced, learning decreased as well, with participants failing to learn both contingencies in parallel. Finally, further reduction in the number of correlations resulted in markedly attenuated learning. Although the ability to classify items in multiple ways has been demonstrated in adults (Ross & Murphy, 1999; see also Shafto, Kemp, Mansinghka, & Tenenbaum, 2011, for computational modeling of cross-classification), children (Nguyen, 2007), and infants (Ellis & Oakes, 2006; Mareschal & Tan, 2007), the presented findings are the first to our knowledge to demonstrate the role of correlated structure in learning of multiple mutually-exclusive categories in infancy. These novel findings have important implications for theories of category learning in infancy and beyond. In addition, as we discuss below, the reported findings may have important broader implications, in particular, for the study of cognitive flexibility and memory development.

Research on category learning demonstrated that when learning is supervised, categories that have a single relevant dimension are easier to learn than categories that have multiple dimensions (Ashby et al., 1998: Kloos & Sloutsky, 2008; Kruschke, 1996). This is because attention is allocated (from a single pool) to dimensions or cues, and as a result, redundant dimensions compete for attention (see Kruschke, 1996, for a review). While supervised learning can be fast and efficient, it has some appreciable costs. Most importantly, supervised learning is likely to engage the selection-based system, and there is a consequence of selectivity: selective attention to few category-relevant dimensions results in learned inattention to irrelevant dimensions (see Hoffman & Rehder, 2010, for eye tracking evidence). Such learned inattention may be acceptable if categories were learned one at a time, but it poses a problem for learning of multiple categories in parallel (see Hoffman & Rehder, 2010, for similar arguments).

In contrast, when learning is unsupervised, it is easier to learn categories that have multiple correlated features than those that are based on a single feature (Billman & Knutson, 1996; Kloos & Sloutsky, 2008; Thiessen & Saffran, 2003), and it has been argued that such learning does not require the involvement of the selection-based system (e.g., Sloutsky, 2010). Research presented here provides support for this idea: infants learned multiple dimension-context correlations, and they did so only when there were multiple context features supporting the target dimension. These findings suggest that learning was achieved without the involvement of selective attention: if learning were based on selective attention, (a) there might be costs of selectivity in the form of learned inattention and (b) few context-dimension correlations would have resulted in successful learning.

A number of ideas have been proposed to explain able learning of correlated structures. One possibility is the idea of "focused sampling" (e.g., Billman & Heit, 1988; Bower & Trabasso, 1963; Billman & Knutson, 1996) – the idea that participants sample one dimension at a time and they are more likely to correctly select a relevant dimension when there are many relevant dimensions than when there are few. Another idea is that inter-correlated features facilitate learning by mutually reinforcing each other (e.g., coherent co-variation effect discussed by Rogers & McClelland, 2004). The principal difference between these two ideas is that the former idea is based on selectivity and may have a hard time explaining flexible learning because selective attention to one dimension results in learned inattention other dimensions. In contrast, the latter idea is not based on selectivity, and, as a result, may explain learning of multiple structures in parallel. Furthermore, the fact that redundant structures are learned particularly well when learning is not supervised (Kloos & Sloutsky, 2008) seems to support the latter idea. This latter idea is also consistent with the proposal of multiple systems of category learning – an evolutionary primitive system underlying learning of redundant structures and not requiring selective attention and a later developing system of focusing on one dimension at a time and requiring selective attention (cf., Ashby et al., 1998; see also Love & Gureckis, 2007; Sloutsky, 2010, for reviews).

Given the overall pattern in the reported experiments, we believe that infants learned two sets of correlated features, with items matching in shape in one context and items matching in color in a different context. However, it is also possible that infants simultaneously learned the two categories by learning different "values" of a single dimension in different contexts (e.g., items match in shape in context 1 and do not match in shape in context 2). The current study cannot rule out the alternative account, as both strategies would allow for the simultaneous learning of two categories. At the same time, however, this possibility does not undermine the novelty of the current findings. First, when there was sufficient redundancy in the input, infants learned two categories (Experiments 1 and 2B), without exhibiting costs of selective attention. Slightly decreasing the number of correlated context features was associated with a decrease in flexibility (i.e., infants in Experiment 2 only learned one category), whereas substantially decreasing the number of context features in Experiment 3 resulted in no learning.

And second, even if infants learned a single dimension, they exhibited (at least by developmental standards) remarkable flexibility. Consider the task that is a benchmark of cognitive flexibility – Dimensional Chance Card Sort (DCCS) task (Zelazo et al., 1996). In this task children are presented with two bi-dimensional items – Shape 1/Color 1 and Shape 2/Color 2. Then they are given a set of test items (Shape 1/Color 2 and Shape 2/Color 1), one at a time, and asked to sort these items according to a rule (e.g., to sort by shape). After 6 trials, the sorting rule changes (e.g., if they sorted by shape, they will have to change to color). The major finding is that 3-to-4-year-olds failed the switch and continued sorting by the old rule. In contrast, infants in the reported experiments exhibited evidence of flexibility, but only when there were redundant context cues. Therefore, redundancy may help infants to be flexible, in situations where older children fail.

The reported research may also have important implications for the study of memory development. It has been argued that early memory is often bound to a particular context. For example, Robinson and Pascalis (2004) found that only 18- and 24-month-olds (but not 6- and 12-month-olds) exhibited flexible recognition of the familiarization picture in a new context, whereas recognition in younger infants appeared to be bound to the studied context. In contrast to these studies where the familiarization items were presented on either same or new background, in our studies familiarization items were presented on either same or switched context. To notice the switch, participants had to encode the item- background relation – the ability that exhibits a relatively late onset (but see, Richmond & Nelson, 2009 for suggestive evidence that younger infants can exhibit evidence of relational memory). The fact that 14-month-olds recognized the switch (but only when there were multiple context-item correlations) suggests (a) in contrast to what might be suggested by Robinson and Pascalis (2004), item and context information are not bound by default in infancy, and (b) multiple correlations may facilitate item-context binding. Additional research is needed to further examine both possibilities.

In summary, the reported results indicate that infants can learn multiple contingencies in parallel, without exhibiting costs of selective attention. The results also provide evidence that such learning requires multiple supporting context features. Taken together these findings indicate that the observed learning did not engage selective attention. These results contribute to understanding of category learning in infancy and they have implications for understanding memory and cognitive flexibility early in development.

Acknowledgements

This research is supported by the NSF Grant BCS-0720135 to VMS and by NIH Grants R03HD055527 to CWR and R01HD056105 to VMS.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442–481.

Billman, D., & Heit, E. (1988). Observational learning from internal feedback: A simulation of an adaptive learning method. *Cognitive Science, 12*, 587–625.

Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 458–475.

Bower, G., & Trabasso, T. (1963). Reversals prior to solution in concept identification. *Journal of Experimental Psychology, 21*, 1–14.

Bruner, J. S., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New York: Wiley.

Casey, M. B. (1979). Color versus form discrimination learning in one-year-old infants. *Developmental Psychology, 15*, 341–343.

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on A-not-B. *Child Development, 56*, 868–883.

Ellis, A. E., & Oakes, L. M. (2006). Infants flexibly use different dimensions to categorize objects. *Developmental Psychology, 42*, 1000–1011.

Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: Towards a unified developmental approach. *Developmental Science, 13*, 622–635.

Hoffman, A. B., & Rehder, B. (2010). The costs of supervised classification: The Effect of learning task on conceptual flexibility. *Journal of Experimental Psychology: General, 139*, 319–340.

Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General, 137*, 52–72.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22–44.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science, 8*, 225–247.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned attention. *Psychonomic Bulletin & Review, 7*, 636–645.

Love, B. C., & Gureckis, T. M. (2007). Models in search of a brain. *Cognitive, Affective & Behavioral Neuroscience, 7*, 90–108.

Macario, J. F. (1991). Young children's use of color in classification: Foods and canonically colored objects. *Cognitive Development, 6*, 17–46.

Mareschal, D., & Tan, S. H. (2007). Flexible and context-dependent categorization by 18-month-olds. *Child Development, 78*, 19–37.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207–238.

Nguyen, S. P. (2007). Cross-classification and category representation in children's concepts. *Developmental Psychology, 43*, 719–731.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

Opfer, J. E., & Bulloch, M. J. (2007). Causal relations drive young children's induction, naming, and categorization. *Cognition, 105*, 207–217.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception, 22*, 463–475.

Rakison, D. H., & Poulin-Dubois, D. (2002). You go this way and I'll go that way: Developmental changes in infants' detection of correlations among static and dynamic features in motion events. *Child Development, 73*, 682–699.

Richmond, J., & Nelson, C. A. (2009). Relational memory during infancy: Evidence from eye tracking. *Developmental Science, 12*(4), 549–556.

Robinson, A., & Pascalis, O. (2004). Development of flexible visual recognition memory in human infants. *Developmental Science, 7*, 527–533.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38*, 495–553.

Shafto, P., Kemp, C., Mansinghka, V. K., & Tenenbaum, J. B. (2011). A probabilistic model of cross-categorization. *Cognition, 120*, 1–25.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*, 13.

Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development, 79*, 639–651.

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science, 34*, 1244–1286.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Statistical and stress cues in infant word segmentation. *Developmental Psychology, 39*, 706–716.

Younger, B. A., & Cohen, L. B. (1983). Infant perception of correlations among attributes. *Child Development, 54*, 858–867.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37–63.