Journal of Experimental Child Psychology 115 (2013) 150–162

Contents lists available at SciVerse ScienceDirect

Journal of Experimental Child Psychology

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

Same, varied, or both? Contextual support aids young children in generalizing category labels

Elizabeth R. Goldenberg *, Catherine M. Sandhofer

Department of Psychology, University of California, Los Angeles, CA 90095, USA

ARTICLE INFO

Article history:
Received 16 May 2012
Revised 14 November 2012
Available online 1 March 2013

Keywords:
Context
Generalization
Word learning
Cognitive Development
Aggregate
Decontextualize

ABSTRACT

Children have a difficult time in generalizing among changes in background context. We examined the role of two processes that may aid in generalizing category labels in new contexts. In this study, 2-year-old children were taught novel object categories in one type of contextual condition and were tested for category generalization in a new context. In Experiment 1, children ($N = 48$) learned in one of three conditions: (a) all category instances presented in the same context, (b) all category instances presented in varied contexts, or (c) some category instances presented in the same context and some presented in varied contexts. In Experiment 2, children ($N = 48$) learned in one of three conditions, all of which included presentations in the same context and varied contexts but differed in order. Results from both experiments revealed that children were significantly more likely to choose the correct object when training was in both same and varied contexts regardless of order. The results suggest that contextual factors, by providing both support for aggregation and support for decontextualization, contribute to word learning and generalization for novice word learners.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Young children exhibit considerable competency—yet surprising fragility—in word generalization tasks. Across multiple studies, children readily, appropriately, and seemingly effortlessly generalize new object labels to similar objects (Klibanoff & Waxman, 2000; Samuelson & Smith, 2000; Smith, Jones, Landau, Gershkoff-Stowe & Samuelson, 2002; Soja, Carey, & Spelke, 1991). For example, if chil-

* Corresponding author.
E-mail address: ergoldenb@ucla.edu (E.R. Goldenberg).

0022-0965/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jecp.2012.11.011

dren are given a novel object and told, ‘‘It’s a pestle,’’ they generalize the label to new pestles of the same shape even if the new pestles differ in size, color, and material from the original one. However, with small changes to the background context, children’s competency at extending and generalizing category labels collapses. For example, when young children are asked to generalize a category label in a new context, their performance drops to levels that are no different from chance (Vlach & Sandhofer, 2011). The role of context in generalizing object labels is important to word learning because children are required to generalize object labels in new contexts regularly in their everyday lives. For example, a child might be asked to generalize the label ‘‘toothbrush’’ to a new toothbrush in a never before seen context, such as a toothbrush sitting in a suitcase.

Aggregation

Young children fail to generalize object labels to similar objects if the context differs between training and testing (Vlach & Sandhofer, 2011). One reason for this difficulty may be that category learning requires the learner to recognize the specific properties that matter for a particular category. To give one example, for the category ‘‘spoon,’’ shape matters for category membership, but color and material do not. Aggregating across multiple instances allows the learner to recognize similarities between disparate objects from the same category (Gentner & Namy, 1999; see also Tversky, 1977). For example, combining multiple memories of spoons (e.g., big wooden spoons, little spoons, plastic spoons, metal spoons) allows the relevant and similar features of spoons (i.e., spoon shape) to emerge. By aggregating instances of the experienced objects together, the relevant properties become stronger and the irrelevant features become weaker.

Covarying features (i.e., redundant correlated cues) support aggregating objects together across different instances (Dueker & Needham, 2005; Smith & Yu, 2008; Thiessen & Saffran, 2003; Yoshida & Smith, 2005). For example, in Yoshida and Smith’s (2005) study, Japanese children’s generalization performance increased when perceptual cues were systematically paired with linguistic categories as a covarying feature. Unlike English, Japanese makes no syntactic distinction between count and mass nouns. Japanese children typically use perceptual (not syntactic) information to appropriately generalize mass nouns to other things that are made of the same material and count nouns to other things that are the same shape. Yoshida & Smith (2005) presented Japanese children with a novel sentence frame that covaried with the object’s count or mass status. Even though the syntactic frames were redundant to the perceptual information and were not part of children’s learning history, the children who heard redundant syntactic frames were better at generalizing the novel label to similar objects than children who were not presented with the syntactic frame. Furthermore, in McDonald and Plauche’s (1995) study, children learning the grammatical structure of an artificial language more accurately identified word class when provided with salient correlated cues than when provided with a single cue. Thus, the presence of covarying features aids in the process of aggregating across instances, which is important in generalization.

Because context can serve as a covarying feature, shared contextual information should aid the learner in noticing the similarities between instances (Rothkopf, Fisher, & Billington, 1982). For example, presenting all instances of a category in the same context provides an extra similarity cue. Evidence for the benefit of redundant context between instances comes from Vlach and Sandhofer’s (2011) study, in which children’s generalization performance increased when provided with redundant support from contextual cues. Specifically, children were more likely to generalize a category label to similar objects when provided with redundant contextual cues (the object label and repetitive background context) than when provided with a single cue (the object label). The shortcoming of a single redundant cue for aggregating object similarities suggests that children need—at least early on—additional types of cues to support aggregation.

Although learning from multiple exemplars in the same context may support attention to similarities between otherwise disparate examples, it also may engender contextual dependency. Research in learning and memory has long described the effects of context dependency where recall is lower when the information is recalled in a different context from that in which it was learned and higher when the information is recalled in the same context in which it was learned (e.g., Amabile &

Rovee-Collier, 1991; Borovsky & Rovee-Collier, 1990; Godden & Baddeley, 1975; Hartshorn et al., 1998; Hayne, Boniface, & Barr, 2000; Hayne, MacDonald, & Barr, 1997; Learmonth, Lamberth, & Rovee-Collier, 2004; Robinson & Pascalis, 2004; Rovee-Collier & Dufault, 1991; Rovee-Collier, Griesler, & Early, 1985; Smith, 1982; Smith, Glenberg, & Bjork, 1978; Suss, Gaylord, & Fagen, 2012). For example, 3- and 6-month-old children had lower long-term retention when training and testing occur in different contexts (different colored and patterned bumpers in the infant's crib) than when testing and training occur in the same context (Amabile & Rovee-Collier, 1991; Borovsky & Rovee-Collier, 1990; Rovee-Collier & Dufault, 1991). Similarly, in a word learning study, 2- and 3-year-olds failed to generalize a category label when the context changed between training and testing (Vlach & Sandhofer, 2011). Although the redundant contextual cues provided during training may have aided in aggregating when the cues were present, redundant correlated cues may have also created context dependency.

Decontextualization

Learning in varied contexts aids recall in new contexts (Amabile & Rovee-Collier, 1991; Rovee-Collier & Dufault, 1991; Smith & Vela, 2001; Smith et al., 1978). For example, when adults were asked to recall information in a different room from the one in which the information was learned, performance was higher when information was learned in varied rooms than when information was learned in a single room (Smith, 1982). Similarly, when 3- and 6-month-olds were trained in varied contexts (i.e., crib bumpers with different colors and patterns), long-term retention was higher in a new context than when training took place in one context (only one colored and patterned crib bumper) (Amabile & Rovee-Collier, 1991; Rovee-Collier & Dufault, 1991). This would suggest that if children learn category labels in varied contexts, they should successfully generalize new category instances when tested in a new context. However, this is not the case. In a previous study, 2-year-olds failed to generalize category labels to new objects when they were taught object names in varied contexts and tested in a new context (Vlach & Sandhofer, 2011). Vlach and Sandhofer (2011) presented children with three novel objects that shared the same shape and labeled the objects with a novel noun (e.g., "Look, a wug!"). When each of the three exemplars was presented and labeled in a different context, children failed to extend the novel noun to new objects that shared the same shape, in a new context. Thus, learning in varied contexts still falls short of what is needed for young children to learn words.

Past research suggests that young children have difficulty in generalizing a new category label when learning takes place *either* in only one context *or* in only different non-overlapping contexts. However, we know that repeated learning in the same context helps children to aggregate similar instances in memory, whereas learning in varied contexts helps children to decontextualize object exemplars. In the current study, we examined whether learning in *both* the same and varied contexts will help young children to overcome the difficulty of generalizing a novel label in a new context.

The current study

Altogether, the results of previous studies suggest that children have difficulty in generalizing category labels when small changes are made to the background context between training and testing. Two processes that may support generalization of category labels are aggregating across instances and decontextualization. In two experiments, we examined how context change affects generalization by manipulating the type of contextual support provided to children during training. We provided children with training in contexts that either provided support for aggregation (i.e., the same contexts), provided support for decontextualization (i.e., varied contexts), or provided support for both (i.e., both the same and varied contexts). We propose that 2-year-olds, as novice word learners, require two types of support to generalize in new contexts: support to aggregate *and* support to decontextualize. When provided *only* with support to aggregate, children should notice similarities among different category exemplars but have difficulty in separating the object and label from the background context. When provided *only* with support for decontextualization, children should separate the object label from the background context but have difficulty in grouping exemplars together based on similarities, leading to decreased generalization. We expected children to overcome the difficulty of generalization when given support for *both* aggregating and decontextualizing. That is, by receiving support for both aggregation and decontextualization, children should both notice the similarities between instances and separate them from the background context, enabling generalization.

Experiment 1

Method

Participants

In this experiment, 48 2-year-olds (24 boys and 24 girls, M_{age} = 24.36 months, SD = 2.43) were randomly assigned to one of three conditions: same context, varied context, or interleaved context. This age group was selected because in previous studies children of this age had difficulty in generalizing among changes in background context. Participants were recruited through birth records and preschools in an urban area. An additional 2 children did not complete the experiment (1 due to fussiness and 1 due to experimental error).

To make sure that there were no differences in children's vocabulary among conditions, parents filled out the MacArthur–Bates Communicative Development Inventory: Words and Sentences (MCDI; Fenson et al., 1994). There were no differences in productive vocabulary among conditions, $F(2,41) = 0.438$, $p > .05$. From the three different conditions, 5 children did not return a complete MCDI and were excluded from this analysis.

Design

This experiment used one between-participants variable, which was contextual condition. All three contextual conditions differed during the training phase but were identical during the distractor and test phases. Context was defined as the color and pattern of the large piece of fabric on which the novel objects rested while they were labeled. Table 1 presents examples of all three conditions. In the *same context* condition, the five training presentations were on pieces of fabric with the same color and pattern (i.e., Context A). In the *varied context* condition, the five training presentations were on pieces of fabric with different colors and patterns (i.e., Context A, Context B, Context C, Context D, and Context E). In the *interleaved context* condition, the first, third, and fifth target exemplars were on pieces of fabric with the same color and pattern, whereas the second and fourth exemplars were on pieces of fabric with different colors and patterns (i.e., Context A, Context B, Context A, Context C, and Context A). In all conditions, the distractor and testing phases each took place on a never before seen fabric.

Each child was tested for object label generalization in eight categories. During the training phase, five novel objects and one label (e.g., ''wug''), for each category, were presented to the child over five presentations. All five target exemplars in each category were shape matches but differed in color and texture. During the distractor phase, the child was presented with a novel distractor object that was

Table 1
Contexts presented by condition.

Condition	Training phase					Distractor phase	Test phase
	Target 1	Target 2	Target 3	Target 4	Target 5		
Experiment 1							
Same	Context A	Context A	Context A	Context A	Context A	Context B	Context C
Varied	Context A	Context B	Context C	Context D	Context E	Context F	Context G
Interleaved	Context A	Context B	Context A	Context C	Context A	Context D	Context E
Experiment 2							
Aggregate first	Context A	Context A	Context A	Context B	Context C	Context D	Context E
Aggregate last	Context A	Context B	Context C	Context C	Context C	Context D	Context E
Interleaved	Context A	Context B	Context A	Context C	Context A	Context D	Context E

not labeled. During the test phase, the child was given a forced-choice test to assess generalization of the novel category label.

Materials

Novel objects were constructed out of arts and crafts supplies. All objects were intended to be new to the child (see Appendix). Each category consisted of six exemplars (five used during the training phase and one used during the testing phase). Large (21 × 26-inch) pieces of colorful patterned fabric served as the "context" in each object presentation. All objects, object label pairs, and fabrics were randomized and counterbalanced within and between participants to ensure that performance differences were not due to particular fabric patterns or object shapes.

Procedure

During the training phase, for each of the eight novel categories, the child was presented with five successive training presentations. In each presentation, a target exemplar was presented for a total of 10 s and was labeled two times with a novel word (e.g., "This is the toma. See the toma?"). The eight novel words used in the study were as follows: dax, wug, toma, blicket, fop, gipple, modi and riff.

The distractor phase took place immediately after the training phase. During this phase, the child was presented with a distractor object for 50 s in a novel context (a piece of fabric with a new color and pattern) while the experimenter brought attention to it without labeling it (e.g., "Look at this!"). The function of the distractor object was to provide a test choice option that the child had recently seen other than the target object. The distractor object ensured that a child did not choose the target object during test based on familiarity or recency of exposure.

During the test phase, the child was presented with four choices: a target exemplar, the distractor object, an unfamiliar object, and a familiar object. An unfamiliar object and a familiar object were included in the test to make sure that the child could differentiate among the correctly labeled object (target), a never before labeled object (unfamiliar), and an object for which the child already had a label such as a toy duck (familiar).

Once the child had a chance to touch all of the objects, the experimenter asked the child to retrieve the target object using the target name (e.g., "Where is the toma?"). If the child did not respond to the first request, the request was repeated. When the child handed the experimenter an object, the experimenter recorded the child's response and the next trial began. All three phases were repeated over eight separate trials with entirely new objects, labels, and context fabrics. A child's generalization score was calculated by summing the number of times the child chose the target exemplar at test. Each child's score had a maximum of 8 because each child completed eight trials.

All object presentations took place on top of a piece of colored and patterned fabric. The fabric was folded around the object with the colored side facing in and closed with a rubber band to create the appearance of a bag. The experimenter retrieved the pre-assigned bag on a given presentation and opened the bag on the table in front of the child to reveal the object already positioned on top of the context fabric. This procedure was used to disguise any directly intended context change by the experimenter and further present context as an incidental factor.

The fabric was always in view underneath the object. The child was allowed to pick up the object; however, the child was not permitted to move the object away from the fabric. The first experimenter monitored the position of the object and made sure that the fabric always remained behind the object. This was done to ensure that the context was always paired with the object.

The first experimenter presented all object presentations to the child. A second experimenter was present throughout the entire experiment. The second experimenter monitored the child's attention and ensured that the objects that were to be presented by the first experimenter were in close reach to her.

Results and discussion

The primary question addressed by analyses was whether children scored differently in the three contextual conditions. As Fig. 1 shows, children in the interleaved context condition scored

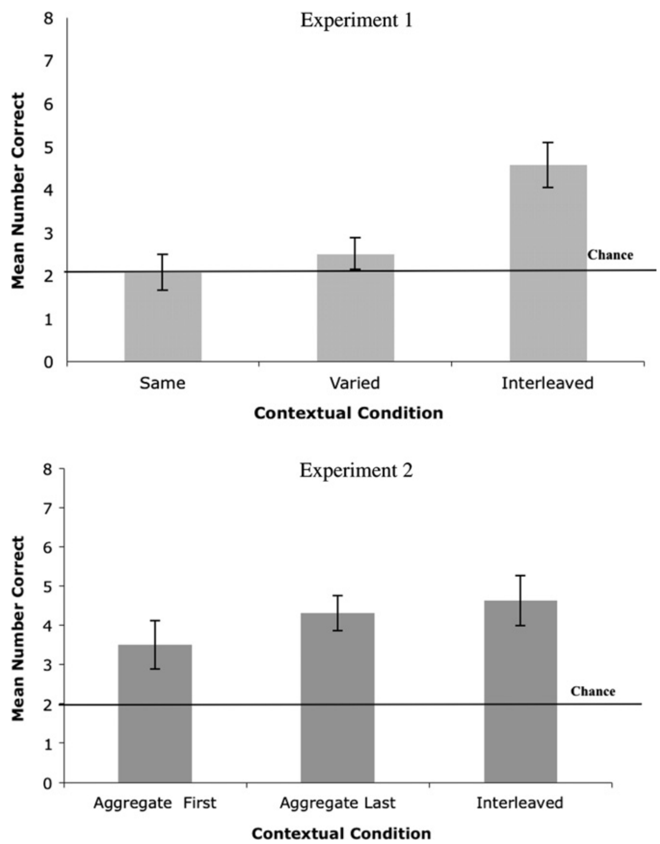

Fig. 1. Results graphed by experiment. Within each experiment, the graph depicts the mean number of correct responses (maximum = 8) by condition. The error bars indicate standard error.

significantly higher than children in the same context condition or the varied context condition. A one-way analysis of variance (ANOVA) revealed a significant main effect of context, $F(2,45) = 9.110$, $p < .001$, $\eta_p^2 = .248$. Planned comparisons, using a Bonferroni correction, revealed that the generalization scores in the interleaved context condition ($M = 4.56$, $SD = 2.09$) were significantly higher than in the varied context condition ($M = 2.50$, $SD = 1.67$), $p = .005$, $d = 1.09$, and significantly higher than in the same context condition ($M = 2.06$, $SD = 1.48$), $p = .001$, $d = 1.37$. There was no significant difference between performance in the varied context condition and performance in the same context condition, $p = .765$.

We also asked whether the generalization scores in each contextual condition were different from chance. We conservatively defined chance as 25% because there were four possible choices for children to select among during the testing phase (children were also allowed to not make a choice by indicating they did not know). Children in the interleaved context condition performed significantly better than chance levels, as revealed by a one-sample *t* test with a test value of 2, $t(15) = 2.562$, $p < .001$, $d = 1.22$. However, children did not perform significantly better than chance in the varied context condition, $t(15) = 0.500$, $p = .251$, or in the same context condition, $t(15) = 0.0625$, $p = .868$.

Across all conditions, children were most likely to choose the target object (39%). When children made an error, they were most likely to choose the distractor object (21%) followed by the unfamiliar object (17%) and the familiar object (14%). In addition, in 9% of the trials, children did not make a

Table 2
Objects selected in test by condition.

Condition	Target (%)	Errors (%)			
		Distractor	Unfamiliar	Familiar	No choice
Experiment 1					
Same	25.78	27.34	19.53	13.28	14.06
Varied	31.25	20.31	20.31	21.86	6.25
Interleaved	57.03	14.84	13.28	7.03	7.81
Experiment 2					
Aggregate first	43.74	23.44	21.88	7.03	3.90
Aggregate last	53.91	23.44	11.72	6.25	4.69
Interleaved	57.81	14.84	14.06	1.56	11.71

choice and indicated that they did not know. As Table 2 shows, there were no systematic differences in children's errors beween conditions.

There were no apparent differences in performance across trials. A paired samples t test revealed no differences in performance across conditions between the first four trials and the last four trials, $t(47) = 1.706$, $p > .05$. Furthermore, there were no differences between performance in the first four trials and performance in the last four trials within either the same context condition, $t(15) = 1.33$, $p > .05$, the varied context condition, $t(15) = 0.00$, $p > .05$, or the interleaved context condition, $t(15) = 1.51$, $p > .05$.

The only difference among the three conditions was the patterned background. Thus, we interpret these findings as evidence for differences in the composition of contextual support during training. The highest generalization performance occurred when two types of contextual support were provided during training, namely (a) support to aggregate, which is provided by training in the same contexts, and (b) support to decontextualize, which is provided by training in varied contexts.

However, another possibility is that children's success in the interleaved context condition had less to do with contextual support and more to do with the interleaving of contexts. In the current experiment, the presence of the same and varied contexts alternated. Thus, the varied contexts were *interleaved* within the same contexts. Research suggests that interleaving can benefit learning and memory (Taylor & Rohrer, 2010) and that interleaving the presentation of categories can benefit generalization (Simon & Bjork, 2001; Taylor & Rohrer, 2010). For example, in one study, Kornell and Bjork (2008) presented participants with paintings categorized by painter. Categories were either presented one at a time or interleaved with other categories. When later asked to categorize new paintings (e.g. "Is this Van Gogh or Monet?"), participants in the interleaved condition scored higher. It is open to question whether interleaving *contexts* benefits generalization in the same way as interleaving *categories.* Thus, perhaps the superiority in the interleaved context condition from Experiment 1 is due to similar interleaving benefits and *not* due to support from aggregation and decontextualization. To test this idea, in Experiment 2 we provided children support to aggregate and decontextualize in both interleaved and non-interleaved conditions.

Experiment 2

In the second experiment, we examined whether the order in which contextual support was provided to learners affected performance in the generalization task. To investigate this, we presented children with support to both aggregate and decontextualize, either by presenting the two types of support interleaved with each other or by presenting one type of support immediately followed by the other type of support. Children received either (a) all of the same contexts *before* the varied contexts, (b) all of the same contexts *after* the varied contexts, or (c) the same contexts interleaved between the varied contexts (as in Experiment 1).

Method

Participants

In this experiment, 48 2-year-olds (27 boys and 21 girls, M_{age} = 25.02 months, SD = 2.82) were randomly assigned to one of three conditions: aggregate first, aggregate last, or interleaved context. Participant recruitment was the same as in Experiment 1. An additional 3 children did not complete the procedure, 1 due to fussiness and 2 due to experimental error.

Design and materials

The materials were the same as in Experiment 1. As Table 1 shows, each condition differed in contextual composition during the training phase. In the *aggregate first* condition, the first three presentations took place on the same type of fabric and the last two presentations were on different types of fabric (i.e., Context A, Context A, Context A, Context B, and Context C). In the *aggregate last* condition, the first two presentations took place on different types of fabric and the last three presentations were on the same type of fabric (i.e., Context A, Context B, Context C, Context C, and Context C). The *interleaved context* condition was the same as in Experiment 1 (i.e., Context A, Context B, Context A, Context C, and Context A).

Procedure

All of the procedures were identical to those in Experiment 1.

Results and discussion

The main question we asked through analyses was whether children were more likely to generalize the novel category label to similar objects in certain contextual orders of presentation. The results revealed that generalization did not differ based on the order of contextual change, $F(2,45) = 1.62$, $p = .208$ (see Fig. 1). The average generalization score was 3.50 (SD = 1.96) for children in the aggregate first condition, 4.31 (SD = 1.40) for children in the aggregate last condition, and 4.63 (SD = 2.02) for children in the interleaved context condition.

We also asked whether the generalization scores in each condition were different from chance (25%). The generalization scores in every condition differed from chance: interleaved context condition, $t(15) = 5.175$, $p < .001$, $d = 1.294$; aggregate first condition, $t(15) = 3.051$, $p = .008$, $d = 0.763$; aggregate last condition, $t(15) = 6.600$, $p < .001$, $d = 1.65$.

Across all conditions, children were most likely to choose the target object (52%). When children made an error, they were most likely to choose the distractor object (20%) followed by the unfamiliar object (16%) and the familiar object (5%). In addition, in 7% of the trials, children did not make a choice and indicated that they did not know. As Table 2 shows, there were no systematic differences in children's errors between conditions.

There were no apparent differences in performance across trials. A paired samples t test revealed no differences between performance in the first four trials and performance in the last four trials across conditions, $t(47) = 0.976$, $p > .05$. The first four trials and last four trials revealed no difference in performance in any of the three conditions: aggregate first condition, $t(15) = 0.355$, $p > .05$; aggregate last condition, $t(15) = 0.745$, $p > .05$; interleaved context condition, $t(15) = 0.565$, $p > .05$.

Children's generalization performance did not appear to be affected by the order of contextual change during training. This suggests that the results from Experiment 1 were not simply due to effects of interleaving support to aggregate with support to decontextualize. Rather, generalization performance is enhanced when both support for aggregation and support for decontextualization are available regardless of presentation order.

General discussion

In the two current experiments, we asked whether children could generalize in the presence of a new background context when provided with different types of support. Across both experiments, we found that specific types of contextual support increase children's generalization performance. In Experiment 1, children in the same context and varied context conditions performed at chance levels. The same context condition provided support to aggregate among instances. The redundant context cues in the same context condition increased the level of similarity between the disparate experiences and, thus, increased the likelihood that the instances would be aggregated together in memory. At the same time, however, the same context condition failed to support decontextualization. In contrast, the varied context condition provided support for decontextualization. Because the contexts changed with each instance, context was marked as a nondefining aspect of the instance. The varying contexts aided in separating the irrelevant contextual features from the relevant category information. However, because the varied context condition lacked aggregative cues, the different instances of the category had no strong cue to link them together. In the interleaved context condition, the presence of both types of support resulted in successful generalization. Importantly, in Experiment 2, where all children were provided with both types of support, children performed above chance levels. These results suggest that the availability of both types of support (support to both aggregate and decontextualize) benefits generalization. Furthermore, this held true with all three of the conditions in Experiment 2, which differed in how the two types of support were ordered.

However, the results of Experiment 2 are contrary to those of studies in which the interleaved nature of presentation increases category induction (Kornell & Bjork, 2008). This may be due to a fundamental difference between the type of information that is interleaved in the current study and that of past studies. For example, in Kornell and Bjork (2008), one category was interleaved with another category. In the current study, categories were always presented in blocks and the context in which exemplars were presented was interleaved. The category interleaving in Kornell and Bjork's (2008) study created time between each category instance, termed *spacing,* which has been shown to aid learning in numerous domains in many age groups (Baddeley & Longman, 1978; Bjork & Bjork, 2011; Vlach, Sandhofer, & Kornell, 2008; Whitten & Bjork, 1977). In contrast, in the current study, each category instance was presented immediately after the previous one and, thus, was not spaced out in time. The difference in temporal presentation of category instances may account for why the interleaved nature of the contextual support in Experiment 2 did not affect generalization. That is, in the current study, even though the types of contextual support were separated in time, the category instances were not, which may be why no added benefit of interleaving was found.

A number of open questions remain. In the current study, the order of same *and* varied contexts did not affect children's performance in generalizing category labels. This is somewhat surprising given that a number of studies have demonstrated that the order in which different category exemplars are presented can have large effects on what is learned (Elman, 1993; Gentner, Loewenstein, & Hung, 2007; Kotovsky & Gentner, 1996; Sandhofer & Doumas, 2008). For example, progressive alignment (Gentner & Markman, 1997; Gentner & Namy, 2006; Gentner et al., 2007; Kotovsky & Gentner, 1996) is a process by which the similarity between exemplars starts out high and decreases as subsequent exemplars are presented. Using this type of method, children are able to generalize at a higher level than when the exemplars are randomly ordered. In the current study, we selected contexts that were obviously different from each other (i.e., little similarity between the patterned fabrics). Perhaps ordering effects would be more likely in instances where the contexts gradually became less similar. A second possibility as to why we did not observe order effects involves the number and ratio of presentations in the same context and varied contexts during training. In Experiment 2 the ratio of same to varied contexts always remained constant (i.e., three contexts that were the same as each other and two contexts that were different from each other). Because a number of studies have suggested that complex learning problems may benefit from reducing the scope of the learning problem, such as in Newport's (1990) "less is more" hypothesis and Elman's (1993) "starting small" proposal, children may profit from factors that support one aspect of word learning (e.g., aggregation) before being exposed to the other aspects (e.g., decontextualization). This might suggest that different numbers and ratios of presentations (e.g., presenting children with six of the same contexts followed by two different contexts) may lead to different patterns of results. Thus, although we did not find differences in the order of contextual presentations in the current experiments, further research should examine whether different ways of ordering context during learning influences generalization performance.

A second open question involves how context and generalization interact over various time scales and retention intervals (either shorter or longer). In one study, when generalization was assessed after a 15-min delay, adding small amounts of time between presentations of category exemplars resulted in large increases in generalization scores (Vlach, Ankowski, & Sandhofer, 2012). Thus, it is possible that changes in contextual support may show some effects of timing changes. However, there is some suggestion that training in varied contexts provides benefits that last over longer time intervals. Rovee-Collier and Dufault (1991) found that when infants were assessed at retention intervals of 3 or 14 days (with a reminder), infants who were trained in various contexts still performed higher than infants who were trained in only one context. This suggests that effects of contextual support are stable over time. Further research should investigate how context effects can be mitigated or intensified by different time scales.

Another open question is whether this particular benefit of contextual support to category generalization is specific to this age group. Research on infants and children suggests greater contextual constraints on memory retrieval in children than in adults, such that the reliance on context is stronger during the earlier years (Rovee-Collier & Dufault, 1991). In multiple studies, younger participants have had a more difficult time recalling in a new context than older participants (e.g., Hartshorn et al., 1998; Hayne et al., 1997, 2000; Robinson & Pascalis, 2004; Rovee-Collier & Dufault, 1991; Rovee-Collier et al., 1985). For example, imitation in a new context was more difficult for 6-month-olds than for 12-month-olds (Hayne et al., 2000). It is possible that the support to aggregate and decontextualize, which was found to be so useful to 2-year-olds in the current study, would not be necessary for older children, as suggested by studies in which 4-year-olds can generalize in a new context regardless of the training context (Vlach & Sandhofer, 2011). Furthermore, it is possible that children younger than 2 years would still have difficulty in generalizing a category label in a new context even with the support of both the same and varied contexts. However, there is some suggestion that support from same *and* varied contextual cues is beneficial for 5-month-olds in a very different paradigm—that of learning object segmentation (Dueker & Needham, 2005). In Dueker and Needham's (2005) study, infants familiarized to an object by three different people in three different places (varied contexts) did not look longer at an object that violated the segmentation they had learned. However, infants who were familiarized to the object segmentation in three different places by three different people (varied contexts), all of whom wore purple gloves (same context), looked longer at violations of the object segmentation they had learned. Given these developmental findings, future research should explore whether same *and* varied contexts support younger children's ability to generalize category labels.

The current experiments also raise a fundamental question about the relationship between generalization and memory. Memory is a critical component in word learning. Learning words requires remembering the link between a word and a particular instance, binding new instances with memories of previous instances, and retrieving words when needed. A long history of research has focused on the factors that promote or hinder memory. The memory tasks in these experiments critically differ from word learning tasks in that memory tasks uniformly examine recall or recognition of exact information (e.g., lists of words). Less is known about how these factors promote or hinder learning in situations where learners need to generalize beyond the information given, such as when children need to generalize a label to a new object in a different context. Because of the differences between memory tasks and word learning tasks, similar situations may differentially affect performance in memory tasks versus generalization tasks. For example, interference may cause subjects to recall the general similarities (Hintzman & Curran, 1994) or gist (Reyna & Brainerd, 1995) common between many instances but fail to remember distinctive item-specific information that distinguishes one episode from another. This is bad for performance in exact memory tasks but characterizes word learning. At the same time, because word learning depends on remembering instances from one moment to the next, the same factors that affect memory should affect word and category learning, and some evidence suggests that categorization may best be described as a memory retrieval processes (Hayne, Rovee-Collier, & Perris, 1987). Future research should examine how context similarly and differentially affects children's performance in memory and generalization tasks. This is a promising route for better understanding the relationship between generalization and memory.

Novice word learners are faced with challenges in categorizing, and we examined one of these—the challenge of generalizing category labels in a completely novel context. The current study aimed to understand why children have such difficulty and what (if any) support from the environment could aid the process. After providing young children with various types of contextual support during training, we conclude that children need two types of contextual support when category generalization is required in a new context. Namely, children need both support to aggregate and support to decontextualize relevant features. Although the current study was experimental in nature, we expect that children are provided naturally with both types of support in everyday environments.

Acknowledgments

We thank Haley Vlach, Mariel Kyger, and Natsuki Atagi for their feedback on previous drafts. We also thank all of the research assistants in the Language and Cognitive Development Lab for their help on this project. We are grateful to the staff, parents, and children at Bunny Patch Preschool, Wagon Wheel Preschool, Attias Daycare Center, the UCLA Infant Development Program, ABC Learning Center, Woodcrest Preschool, Gan Yaffa Preschool, Kids Point Preschool, Wilshire Preschool, Y2Kids Preschool, Tender Loving Care School, and Hill N Dale Preschool.

Appendix. Novel category exemplars

References

Amabile, T. A., & Rovee-Collier, C. (1991). Contextual variation and memory retrieval at 6 months. *Child Development, 62*, 1155–1166.

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics, 21*, 627–635.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York: Worth.

Borovsky, D., & Rovee-Collier, C. (1990). Contextual constraints on memory retrieval at 6 months. *Child Development, 61*, 558–594.

Dueker, G., & Needham, A. (2005). Infants' object category formation and use: Real world context effects on category use in object processing. *Visual Cognition, 12*, 1177–1198.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71–99.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the society for research in child development* (Vol. 59, 5, Serial No. 242).

Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development, 8*, 285–307.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45–56.

Gentner, D., & Namy, L. (1999). Comparison and the development of categories. *Cognitive Development, 14*, 487–513.

Gentner, D., & Namy, L. L. (2006). Analogical processes in language learning. *Current Directions in Psychological Science, 15*, 297–301.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*, 325–331.

Hartshorn, K., Rovee-Collier, C., Gerhardstein, P. C., Bhatt, R. S., Klein, P. J., Aaron, F., et al (1998). Developmental changes in the specificity of memory over the first year of life. *Developmental Psychobiology, 33*, 61–78.

Hayne, H., Boniface, J., & Barr, R. (2000). The development of declarative memory in human infants: Age related changes in deferred imitation. *Behavioral Neuroscience, 114*, 77–83.

Hayne, H., MacDonald, S., & Barr, R. (1997). Developmental changes in the specificity of memory over the second year of life. *Infant Behavior and Development, 20*, 233–245.

Hayne, H., Rovee-Collier, C., & Perris, E. E. (1987). Categorization and memory retrieval by three-month-old infants. *Child Development, 58*, 750–767.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33*, 1–18.

Klibanoff, R. S., & Waxman, S. R. (2000). Basic level object categories support the acquisition of novel adjectives: Evidence from preschool-aged children. *Child Development, 71*, 649–659.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585–592.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*, 2797–2822.

Learmonth, A. E., Lamberth, R., & Rovee-Collier, C. (2004). Generalization of deferred imitation during the first year of life. *Journal of Experimental Child Psychology, 88*, 297–318.

McDonald, J. L., & Plauche, M. (1995). Single and correlated cues in an artificial language learning paradigm. *Language and Speech, 38*, 223–236.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11–28.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*, 1–75.

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

Rothkopf, E. Z., Fisher, D. G., & Billington, M. J. (1982). Effects of spatial context during acquisition on the recall of attributive information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 126–138.

Rovee-Collier, C., & Dufault, D. (1991). Multiple contexts and memory retrieval at 3 months. *Developmental Psychobiology, 24*, 39–49.

Rovee-Collier, C., Griesler, P. C., & Early, L. A. (1985). Contextual determinants of retrieval in three-month-old infants. *Learning and Motivation, 16*, 139–157.

Samuelson, L. K., & Smith, L. B. (2000). Children's attention to rigid and deformable shape in naming and non-naming tasks. *Child Development, 71*, 1555–1570.

Sandhofer, C. M., & Doumas, L. A. A. (2008). Order of presentation effects in learning color categories. *Journal of Cognition and Development, 9*, 194–221.

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 907–912.

Smith, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition, 10*, 405–412.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6*, 342–353.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. K. (2002). Object name learning provides on-the-job training for attention. *Psychological Science, 13*, 13–19.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin and Review, 8*, 203–220.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*, 1558–1568.

Soja, N., Carey, S., & Spelke, E. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition, 38*, 179–211.

Suss, C., Gaylord, S., & Fagen, J. (2012). Odor as a contextual cue in memory reactivation in young infants. *Infant Behavior & Development, 35*, 580–583.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*, 837–848.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of statistical and stress cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39*, 706–716.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.
Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 246–254.
Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children's context-dependent word learning. *Journal of Experimental Child Psychology, 108*, 394–401.
Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition, 109*, 163–167.
Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: The effects of spacing. *Journal of Verbal Learning and Verbal Behavior, 16*, 465–478.
Yoshida, H., & Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science, 16*, 90–95.