



## Journal of Cognition and Development

Publication details, including instructions for authors and subscription information:  
<http://www.informaworld.com/smpp/title~content=t775653659>

### Order of Presentation Effects in Learning Color Categories

Catherine M. Sandhofer <sup>a</sup>; Leonidas A. A. Dumas <sup>b</sup>

<sup>a</sup> University of California, Los Angeles <sup>b</sup> University of Hawaii, Manoa

Online Publication Date: 01 April 2008

**To cite this Article** Sandhofer, Catherine M. and Dumas, Leonidas A. A.(2008)'Order of Presentation Effects in Learning Color Categories',*Journal of Cognition and Development*,9:2,194 – 221

**To link to this Article:** DOI: 10.1080/15248370802022639

**URL:** <http://dx.doi.org/10.1080/15248370802022639>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Order of Presentation Effects in Learning Color Categories

Catherine M. Sandhofer

*University of California, Los Angeles*

Leonidas A. A. Doumas

*University of Hawaii, Manoa*

Two studies, an experimental category learning task and a computational simulation, examined how sequencing training instances to maximize comparison and memory affects category learning. In Study 1, 2-year-old children learned color categories with three training conditions that varied in how categories were distributed throughout training and how similarity between exemplars progressed across instances. The results indicate that beginning learning by interacting with a limited set of highly similar exemplars leads to more learning than when the instances are distributed and dissimilar. In Study 2, order effects were examined with a symbolic connectionist model of general learning and representation discovery (DORA). The results of the studies suggest that when the presentation of instances is ordered in such a way that discrete instances of a category can be more readily connected in memory, category learning and discovery are more likely to occur.

A long history of research on learning categories and concepts has focused on the learning situations that make categories easier versus harder to learn. This line of work has focused on the role of prior knowledge (Kaplan & Murphy, 2000; Pazzani, 1991; Wattenmaker et al., 1986), category labels (Barsalou, 2003; Kersten & Smith, 2002; Kotovsky & Gentner, 1996), and the types and range of instances presented (Posner & Keele, 1968; Ross, 1996, 1997; Yamauchi & Markman, 1998). Considerable research suggests that children and adults learn more quickly and accurately if the learning

---

Correspondence should be sent to Catherine M. Sandhofer, University of California, Los Angeles, 2253B Franz Hall, Box 951563, Los Angeles, CA 90095. E-mail: sandhof@psych.ucla.edu

trials include a full range of instances of the category. For example, in learning the category of "red," a child exclusively presented with instances of red cars would most likely form a very narrow category of redness in which red pencils and red airplanes are rejected as possible category members. Moreover, presenting such a narrow range of instances of the category "red" would perhaps lead the learner to an incorrect hypothesis about the category (e.g., presuming that "red" refers to wheeled things or perhaps things that are red *and* car shaped). If, on the other hand, a learner is presented with a broad range of category instances (e.g., a set that includes red cars, red socks, red apples, red paint, and other diverse red instances), the learner should be able to deduce that red pencils and red airplanes are category members.

In this line of work, the focus has been on showing the types of instances that affect category learning and less consideration has been given to the particular order or timing in which these instances are experienced. However, a second line of work, largely emanating from the language acquisition literature, has focused on the sequence in which specific types of examples are presented during the learning history. This literature began with the somewhat paradoxical hypothesis that starting with less information may result in greater success in the end. The "Less Is More" hypothesis (Newport, 1990) and the incremental learning model (Elman, 1993) have argued that there are advantages in beginning the learning process with a limited set of information.

Elman (1993), for example, simulated this idea by training a neural network model to process complex sentences by presenting the model with sentences in which it received one word at a time and was asked to predict the next word. The network failed to learn when the set of sentences were presented in random order. However, when the network was trained incrementally, starting by processing the simplest sentences from the set and gradually processing the more complex sentences, the network was able to learn the task (but see Rohde & Plaut, 1999). Newport (1990) has used a similar line of reasoning to argue that younger versus older children's increased fluency with language learning may be due to cognitive limitations of the younger learners that effectively create situations in which initially only small amounts of the input are processed. As the child becomes older and the cognitive limitations decrease, more of the input becomes available for processing. There is some empirical support for this idea. Kersten and Earles (2001) found that subjects were better able to learn an artificial language when they were initially presented with only small segments of language than when they were presented immediately with the full complexity of the language. As a whole, these studies suggest that complex problems may sometimes best be solved by ordering the training instances such that

learning begins with a subset of relatively easy examples that have little variation and then increases gradually in complexity and variability.

### ORDERING INSTANCES BY SIMILARITY

One way to limit initial variation is to order training instances such that each instance is highly similar to the next instance. When similarity between instances changes incrementally, it may be easier to notice which attributes matter for category membership and which do not. Support for this idea comes from adult studies in which subjects are given a list of familiar features and are asked to determine which feature or combination of features matter for category membership (Elio & Anderson, 1984; Medin & Bettger, 1994; Wattenmaker, 1993). For example, in Wattenmaker's study, subjects were given index cards each depicting four binary attributes: the person's favorite sport (either tennis or golf), type of hobby (either painting or ceramics), music preference (either jazz or classical), and marital status (either single or married). The subjects are asked to determine which attribute or attributes matter for membership in the "Koala Club" or "Dolphin Club." In these types of experiments, when the cards are ordered such that only one attribute varies at a time (i.e., similarity changes incrementally), subjects are more accurate at finding the critical attributes.

The juxtaposition of highly similar instances may also increase attention to the shared features between successive instances. Work by Gentner and colleagues (Gentner, 1988, 1989; Gentner & Namy, 1999; Gentner & Rattermann, 1991) has shown that when instances share a high level of perceptual similarity, learners are more likely to compare the instances and notice the similarities and differences between them. Indeed Gentner and colleagues (Gentner, 1988, 1989; Gentner & Namy, 1999; Gentner & Rattermann, 1991) have argued that similarity and comparison are the important drivers in the development of knowledge. Initially only instances with massively overlapping literal similarity are likely to be compared (see Gentner, 1988, for review). But through the process of comparison, children begin to develop a knowledge base, which allows children to make less similar matches. And as children gain ever more knowledge, they become better able to make matches without perceptual similarity support.

In one demonstration of this idea, Kotovsky and Gentner (1996) found that the order in which instances were presented affected children's relational matching in a task where children were asked to combine two known dimensions (size and darkness) to discover an unknown higher-order relation. In the study, children saw a target array consisting of a series of circles that increased in size and two choice arrays: a series of squares that

increased in darkness or a set of squares unordered in darkness. Children were asked to choose which of the choice arrays matched the target array. Finding the correct answer required noticing that the squares increasing in darkness and the circles increasing in size both involve the relation of increasing in magnitude. Under normal circumstances, this kind of cross-dimensional matching problem is quite difficult for 4-year-old children. In their study, Kotovsky and Gentner attempted to teach children to find the cross-dimensional match using different training conditions. When the examples were randomly ordered, children responded at chance levels. However, in a training regimen termed "progressive alignment," children were presented with an ordered series of trials that began with mappings based on overall similarity. In this regimen, children were presented with a training series based on simple one-dimensional mappings (i.e., ascending size series mapped to an ascending size series and ascending darkness series mapped to ascending darkness series). In each case, children began with easy holistic matches (e.g., matching dark circles to dark circles) and progressed to harder ones (e.g., matching dark circles to dark squares). Children who received progressive alignment training succeeded in learning the abstract cross-dimensional mappings without feedback and virtually without error. This study provides support for the idea that beginning learning with limited amounts of variation in the initial instances may ease the discovery of difficult and unfamiliar categories.

### DISTRIBUTING INSTANCES

Another line of studies in which ordering of trials affects task performance originates from the memory literature. In this literature the robust finding among adults (e.g., Toppino, 1991), infants (e.g., Vander Linde, Morrongiello, & Rovee-Collier, 1985), and animals (e.g., Fanselow & Tighe, 1988) is that memory for instances is strongest when information is presented over temporally discontinuous sessions instead of presented within a single, massed session (see Ebbinghaus, 1885; Underwood, 1961). In a typical experiment, subjects are asked to study a list of target items that are repeated multiple times during the experiment. The number of intervening trials between the presentations of the target items defines the spacing. For example, Toppino (1991) compared the retention of items in distributed presentations, in which items recurred three times with five other items presented between each reoccurrence, to the retention of items in massed presentations, in which items recurred three times with no other items presented between each reoccurrence. Toppino found that 3- and 4-year-old children recalled

more items when the presentations were distributed than when the presentations were massed.

Although studies like this clearly indicate that the way instances are ordered affects performance, the memory tasks fundamentally differ from tasks in the language acquisition and cross-dimensional mapping studies. Critically, the memory tasks do not require subjects to induce, discover, or generalize, but rather subjects are only asked to remember specific items (e.g., lists of words) over some time period. Thus, the effect of distributing instances in temporally disjunct sessions is likely to have differential effects on performance if the to-be-learned category is familiar or unfamiliar to the subject. Discovering unfamiliar categories requires subjects to remember instances across periods of time, but the memory of a specific item needs to endure only to the point in time in which it could be compared with a new instance.

In this article we examine whether the order in which training instances are presented, and not just the content of what is presented, affects what is ultimately learned. The particular effects we are concerned with in the present studies are not the type of order effects that researchers take great pains to avoid through counterbalancing (such as practice or fatigue effects), but rather deliberate ways of sequencing trials to maximize memory and similarity between instances. We begin by reporting a study of children's category learning in which we vary the initial similarity of category exemplars and the distribution of category instances across training. We next report a network simulation designed to address issues of what differences in performance based on the order in which learning trials are presented may mean for learning more generally.

## STUDY 1

In Study 1 we trained children using three training conditions: (a) Ordered Similarity and Massed Presentation (OM), in which children began with highly similar whole-object matches that progressively became less similar (in this condition, children were trained with massed presentations of one category at a time); (b) Random Similarity and Massed Presentation (RM), in which the similarity between instances was randomly ordered but children were presented with massed presentations of one category at a time; and (c) Ordered Similarity and Distributed Presentation (OD), in which children began with highly similar whole-object matches that progressively became less similar and the category instances were evenly distributed throughout training such that children learned all three categories simultaneously.

The categories we attempted to teach children in this study were categories of color (e.g., red, green, yellow). We chose to train children on this type of category because of children's well-reported difficulty in learning color categories (see Sandhofer & Smith, 1999, for review), and we chose these colors as a contrast to other color training studies. Rice (1980), for example, found that children needed as many as 1,080 trials to learn to correctly label the three colors red, green, and yellow. Throughout the study we use the term *massed* to refer to situations in which one color is taught at a time in massed training blocks and the term *distributed* to refer to situations in which training of different colors is intermixed throughout training. Although this correlates with the timing of presentations, at stake in the present study is the heterogeneity and homogeneity of the training trials.

In all conditions, children are trained using a triad-matching task. In each trial, children are presented with an exemplar, told its color (e.g., "This is red"), and asked which of two choice objects—a target or a distracter—is also red. All children in all conditions experienced the same 18 triad matches, heard the same instructions, and received the same feedback. What differed between the three conditions was only the order in which the triads were presented. We compare children's performance in each of the three conditions in the training trials and in transfer tasks.

## Method

### *Participants*

The participants were 42 monolingual English-speaking 2-year-old children (mean = 28.8 months, range 24.5–34.9,  $SD = 3.9$  months). Children were recruited from preschools in Bloomington, Indiana, and Los Angeles, California. Children were randomly assigned to one of the three training conditions. In each training condition, half of the children were male and half were female. Children were initially screened for color knowledge using a color pretest, and children who made three correct selections (out of 3 questions) were not eligible to participate. In addition, children with a family history of color blindness were not eligible to participate.

### *Procedure and Materials*

Children participated in the tasks in the following order: (a) pretest, (b) training, (c) memory test, and (d) posttest. The entire study took approximately 20 minutes to complete. Children were seated facing the experimenter. Once the child was comfortable, the experimenter asked the

child if they wanted to “play a game with the toys I brought.” Table 1 shows the sequence of events in the experiment.

*Pretest/posttest.* Children were queried on their ability to comprehend the terms *red*, *yellow*, and *green*. Stimuli consisted of one red, one yellow, and one green fabric square. The colors were primary colors easily identifiable by adults as red, green, or yellow. The squares were placed in front of the child in a random sequence, and the child was asked to indicate a particular color, for example, “Show me red.” After the child made a selection, the child was thanked, the squares were reordered and replaced in front of the child, and the next color was queried.

*Training.* In each trial, children were shown one of three exemplar objects, a color matching choice, and a distracter object. The three exemplar objects used in the training trials were a large red cup, a large yellow shoe, and a large green dinosaur. The color matching objects and distracter

TABLE 1  
An Example of the Sequence of Events in the Three Conditions

<i>Ordered Similarity and Massed Presentation (OM)</i>	<i>Ordered Similarity and Distributed (OD)</i>	<i>Random Similarity and Massed (RM)</i>
1. Pretest	1. Pretest	1. Pretest
2. Training	2. Training	2. Training
Whole-object trial (Red)	Whole-object trial (Red)	Attribute trial (Red)
Whole-object trial (Red)	Whole-object trial (Red)	Whole-object trial (Red)
Attribute trial (Red)	Whole-object trial (Yellow)	Attribute trial (Red)
Attribute trial (Red)	Whole-object trial (Yellow)	Attribute trial (Red)
Attribute trial (Red)	Whole-object trial (Green)	Whole-object trial (Red)
Attribute trial (Red)	Whole-object trial (Green)	Attribute trial (Red)
Whole-object trial (Yellow)	Attribute trial (Red)	Attribute trial (Yellow)
Whole-object trial (Yellow)	Attribute trial (Red)	Attribute trial (Yellow)
Attribute trial (Yellow)	Attribute trial (Yellow)	Whole-object trial (Yellow)
Attribute trial (Yellow)	Attribute trial (Yellow)	Attribute trial (Yellow)
Attribute trial (Yellow)	Attribute trial (Green)	Whole-object trial (Yellow)
Attribute trial (Yellow)	Attribute trial (Green)	Attribute trial (Yellow)
Whole-object trial (Green)	Attribute trial (Red)	Attribute trial (Green)
Whole-object trial (Green)	Attribute trial (Red)	Attribute trial (Green)
Attribute trial (Green)	Attribute trial (Yellow)	Attribute trial (Green)
Attribute trial (Green)	Attribute trial (Yellow)	Whole-object trial (Green)
Attribute trial (Green)	Attribute trial (Green)	Attribute trial (Green)
Attribute trial (Green)	Attribute trial (Green)	Whole-object trial (Green)
3. Memory test	3. Memory test	3. Memory test
4. Posttest	4. Posttest	4. Posttest



objects were smaller than the exemplars. The entire set of object triads is listed in the Appendix.

At the beginning of each trial, the child was given the exemplar, color match, and distracter objects to play with for 15–30 seconds. The objects were taken away, and the experimenter held up the exemplar object and said, for example, “This is red. This is red. It goes in here,” and then placed the exemplar in a transparent bucket. The two choice objects were then placed equidistantly in front of the child, and the experimenter said, for example, “Which one’s red? Get the red one.”

If the child selected the correct color matching choice, the experimenter said, “You’re right! Put it in there (*the transparent bucket*). It’s red (*holding up the exemplar object*) and it’s red (*holding up the child’s choice object*). Good for you!”

If the child selected the wrong object, the experimenter took the distracter from the child and placed the color match in the child’s hand while saying, “No, that’s not red. This is red; it goes in here (*the transparent bucket*). It’s red (*holding up the exemplar object*) and it’s red (*holding up the child’s choice object*). Good for you!” Thus, although children may have initially chosen an incorrect color match, they were essentially forced to ultimately select the correct color matching choice. The objects were then removed and the experimenter suggested, “That was fun. Let’s play again!” Children were allowed to work at their own pace; however, children who refused to make a selection after three explicit requests were treated in the same manner as children who had made an incorrect selection.

There were two types of triad matches: whole-object similarity matches and attribute matches. Examples of the triad types are shown in Figure 1. In the *Whole-Object Similarity* trials, children received two whole-object similarity trials for each of the three colors queried. The choice objects in each trial included the color match and the distracter. In addition to matching the exemplar in color, the color matching objects in these trials also matched by object kind, e.g., matching a large yellow shoe (exemplar) to a small yellow shoe (color match). The distracter object did not match the exemplar object in any way, e.g., a red plastic flower. In Figure 1, boxes a, d, and g show examples of whole-object similarity trials. Note that the color match and distracter choice objects were different from each other in object kind and, in that way, highly discriminable from one another. Thus, in these trials the child could select the correct choice merely by choosing the object that was most similar to the exemplar overall.

Children received four Attribute trials for each of the three colors queried. The choice objects in each trial included a color matching object and a distracter object. The color match was identical to the exemplar

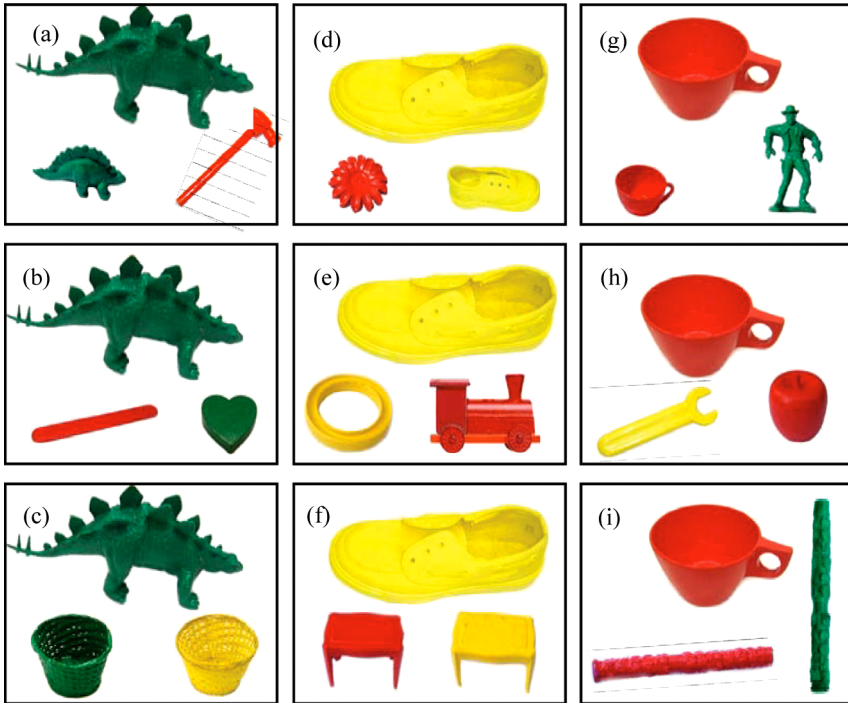


FIGURE 1 Examples of the objects used in the experiment. Whole-object similarity and the two types of attribute trials are shown for each of the three colors.

object in color, but not in other properties, e.g., matching the large red cup (exemplar) to a red apple (color match). The distracter object did not match the exemplar object in any way, e.g., a yellow wrench. In half of the trials, the color match and distracter choice objects were different from each other in object kind, e.g., a red apple (color match) and a yellow wrench (distracter). In Figure 1, boxes b, e, and h show examples of these attribute trials. In the other half of the trials, the color match and distracter choice objects were identical to each other in every way except for color, e.g., a red table (color match) and a green table (distracter), and thus were discriminable only in the attribute of color. In Figure 1, boxes c, f, and i show examples of these trials.

Children participated in one of three training conditions. Table 1 illustrates the sequence of events for the three conditions. The three conditions OM, RM, and OD differed only in the *order* in which the trials were presented; thus all children were presented with all trials.

In the OM condition, children were presented with Whole-Object Similarity matches first and the trials for each of the colors in blocks. That is, all of the training trials for an individual color were presented before the trials for another color were introduced, and the training trials for each color began with whole-object similarity matches. Specifically, for each color, children were first presented with the two whole-object matches followed by the four Attribute matches. Which color block was presented first, second, or third was randomly determined for each child.

In the OD condition, children were presented with Whole-Object Similarity matching trials first, but training trials for the three colors were interspersed (i.e., not blocked). Which color was presented first, second, or third was randomly determined for each child.

In the RM condition, children were presented with all of the trials for each of the colors in blocks, but within each color block, children were presented with the Whole-Object Similarity and Attribute matches in random order. Which color block was presented first, second, or third was randomly determined for each child.

*Memory test.* Children were presented with the three exemplar objects and asked to identify each of the three colors queried, for example, “Show me red.” No feedback was provided.

## Results

### *Pretest*

There was no difference in the pretest scores for children in the three conditions: OM (mean = 1.14,  $SD = .66$ ), RM (mean = 1.14,  $SD = .77$ ), OD (mean = 1.21,  $SD = .58$ ). Thus, prior to training, children in the three conditions did not differ in performance.

### *Training*

Children’s performance on the training trials for the three conditions is shown in Figure 2. As can be seen, children in the OM condition performed significantly better in both the Whole-Object Similarity Matches and the Attribute matches than children in the other two conditions. A 3 (condition)  $\times$  2 (trial type) repeated measures ANOVA revealed a main effect of condition,  $F(2, 39) = 19.49$ ,  $p < .01$ , and a main effect of trial type,  $F(1, 39) = 12.29$ ,  $p < .01$ , but no interaction. Post hoc analyses revealed that children in the OM condition scored significantly higher in the training trials than children in the other two conditions (Tukey’s HSD  $p < .05$ ). Post hoc

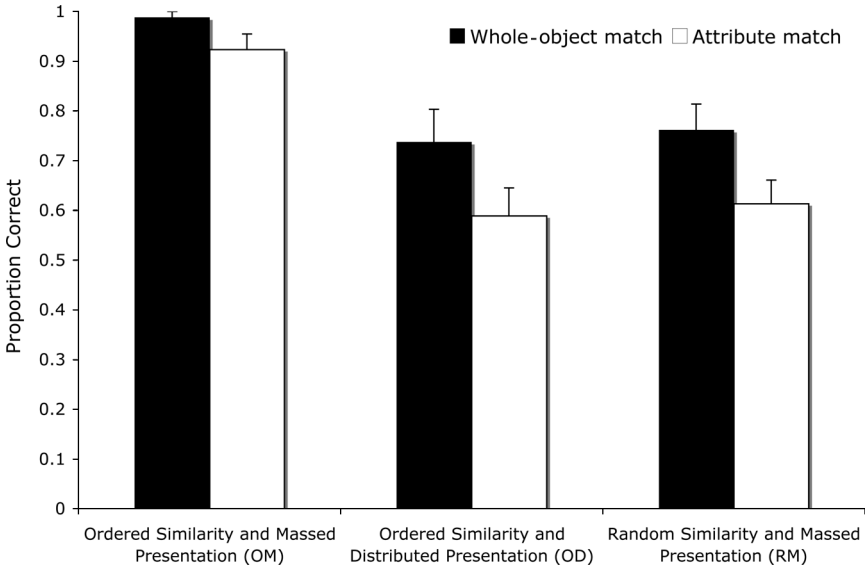


FIGURE 2 Children's performance in the training trials for each of the three conditions.

analyses also revealed that children in all conditions scored significantly better in the whole-object match trial types than in the attribute trials.

### Memory Test

In the memory test, children were shown the three exemplar objects and were asked, for example, "Show me red." Figure 3 shows children's performance in the memory trials. Children in the OM condition and children in the OD condition chose the correctly colored exemplar at levels that exceeded chance performance, OM condition mean = 2.21,  $SD = .80$ ,  $t(13) = 5.67$ ,  $p < .01$ ; OD condition mean = 1.93,  $SD = .127$ ,  $t(13) = 2.74$ ,  $p < .05$ , but children in the RM condition did not (RM mean = 1.43,  $SD = .94$ ).

### Posttest

In the posttest, children were shown the fabric squares from the pretest and were asked, for example, "Show me red." Figure 3 shows the posttest scores in the three conditions. Children in the OM condition showed a significant increase in color comprehension on the posttest compared to their scores on the pretest,  $t(13) = 3.12$ ,  $p < .01$ ,  $d = .83$ , but the children in the other two conditions did not show a significant increase in color comprehension on

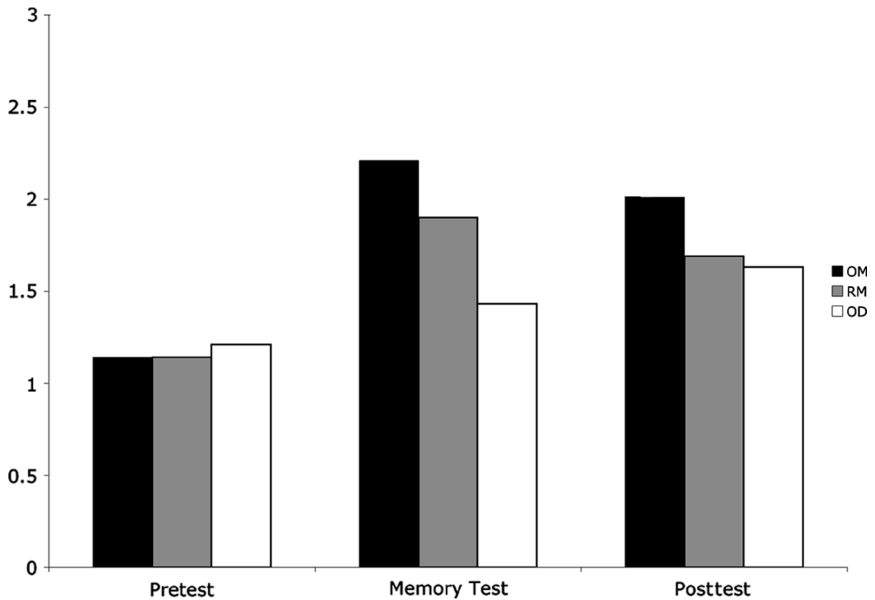


FIGURE 3 Children's performance in the pretest, memory test, and posttest for each of the three conditions.

the posttest: OM mean = 2.00,  $SD = .96$ , RM mean = 1.71,  $SD = 1.07$ , OD mean = 1.64,  $SD = 1.01$ . Thus, not only did children in the OM condition demonstrate that they had learned about color during the 18-trial experiment, but children in the OM condition were able to find the correct color choices even when the choices present the far transfer of the posttest items.

These results show that the order in which instances were presented in a category learning task affected what was learned from the task. Specifically, both beginning by learning labels that applied to holistically similar things and mastering one color word before learning a second color word appear to aid learning. Moreover, the magnitude of this type of training appears to be largely based on the effect size found between pretest and posttest scores in the OM condition. Similar to the children in Kotovsky and Gentner's (1996) study, the children in the OM condition in the present experiment were nearly errorless during the training procedure and moreover showed significant improvement in a far transfer posttest task. Children in the OD condition who began by learning all three labels simultaneously and children in the RM condition who began with holistically dissimilar matches did not succeed in the training or transfer tasks. These results are especially striking in that children in all three conditions experienced the exact same examples during the training procedure.

The *only* difference between the three conditions was the order in which the examples were presented.

Why did children in the OM condition show increased learning during the course of the experiment? One possibility is that beginning category learning by starting with a set of highly similar instances facilitates category discovery. This idea is consistent with a study by Casasola (2005), which finds that limiting the amount of object variability aids infants in acquiring abstract spatial categories. In this study, children who were presented with a smaller set of object pairs (two vs. six) looked longer at new spatial relations than at familiar spatial relations, suggesting the infants who viewed a smaller set of objects had formed a spatial category. Additional support for this idea comes from scaffolding studies in which presenting children with problems of increasing difficulty leads to more learning than when children are solely presented with all easy or all hard problems (Vygotsky, 1978). Similarly, Deák, Ray, and Pick (2002) found that children were better able to induce an abstract function sorting rule in a condition where they experienced four training trials that increased in difficulty than in a condition with two training trials that were of similar difficulty. We further examine the efficacy of the OM condition in Study 2 and return to the question of the primacy of the OM condition in the General Discussion.

## STUDY 2

In Study 2, we examine the generality of the OM condition as a facilitator for category discovery by formally modeling the task with a general network learner. Models are by definition simplifications of the real-world complex processes involved in learning. The value in simplifying and simulating the training orders of Study 1 is that we are able to define and isolate the factors that we expect are important in creating the pattern of results and test whether changes in these factors model the pattern of results seen in the children. In doing so, we attempt to describe the system in terms of the underlying mechanisms in the task structure rather than merely noting the correspondence between the three orders and three patterns of results. The goal of this second study was thus to identify whether the same pattern of results we observed in Study 1 would arise when a general learning model was provided with ordered similarity and massed presentations vs. random ordered or distributed structure presented in Study 1.

The general learning model we use in the present study is DORA (Discovery Of Relations by Analogy; Dumas et al., 2008). DORA is a symbolic connectionist model of learning and representation discovery that has accounted for a wide range of phenomena from the domain of dimension,

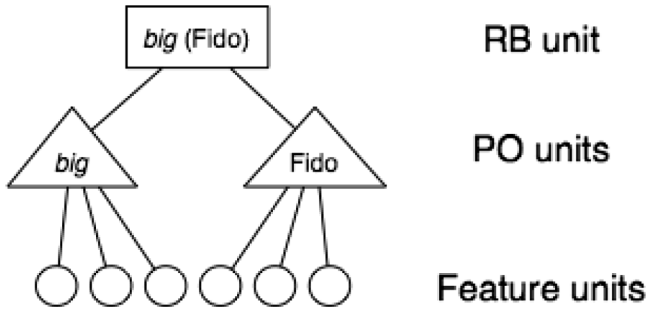


FIGURE 4 A representation of the proposition *big (Fido)* in DORA.

attribute, and relation learning (see Doumas et al., 2008). Importantly, DORA was not created as a means to account for the present data (rather DORA was designed to model relation discovery). Consequently, DORA performance on the task described above serves as a demonstration of how a general learning system performs when exposed to the three order conditions. DORA begins with representations of objects composed holistically as single units (or nodes) connected to unstructured sets of features (Figure 4). Through processes of comparison and intersection discovery (described below), DORA learns explicit (i.e., structured) representations of the properties of compared objects.

In Study 2, we presented DORA with the same training regimens used in Study 1 (i.e., OD, RM, and OM). In the OM condition, the network learned one category at a time, and the stimuli were ordered in such a way that the initial comparisons shared a good deal of overall similarity. In the OD condition, the network was initially presented with comparisons that contained high overall similarity, but the three categories were learned simultaneously. In the RM condition, the network learned one category at a time, but the comparisons within each category were randomly ordered. If the trend observed in Study 1 is a product of the order effects and not specific domain of color, then DORA should perform like the children in Study 1 across all the tested domains.

## Method

### *Architecture*

DORA begins with unstructured representations of objects connected to sets of features (see Figure 4). Eventually, through its learning routines, it comes to represent information in propositional structures using a hierarchy of

distributed and localist codes. We describe the end state of DORA's representations first in order to make the description of DORA's learning routines easier to follow.

At the bottom of the representational hierarchy, *feature* units represent objects and roles (i.e., single-place predicates) in a distributed fashion (e.g., a frog might be represented as "green," "jumps," "has-a-long-tongue," etc., and a clover by "green," "four-leaved," "lucky," etc.; the small circles in Figures 4 and 5). At the next layer of the network, these distributed representations are connected to localist *PO* units that code for specific object or roles. At the next layer, localist *RB* units connect objects and roles into specific role-object pairs (the rectangles in Figures 4 and 5). For example, to explicitly represent that Fido the dog is big (i.e., the property "big" predicated about Fido or the proposition *big* [Fido]), DORA represents Fido as a PO unit connected to the features that define Fido

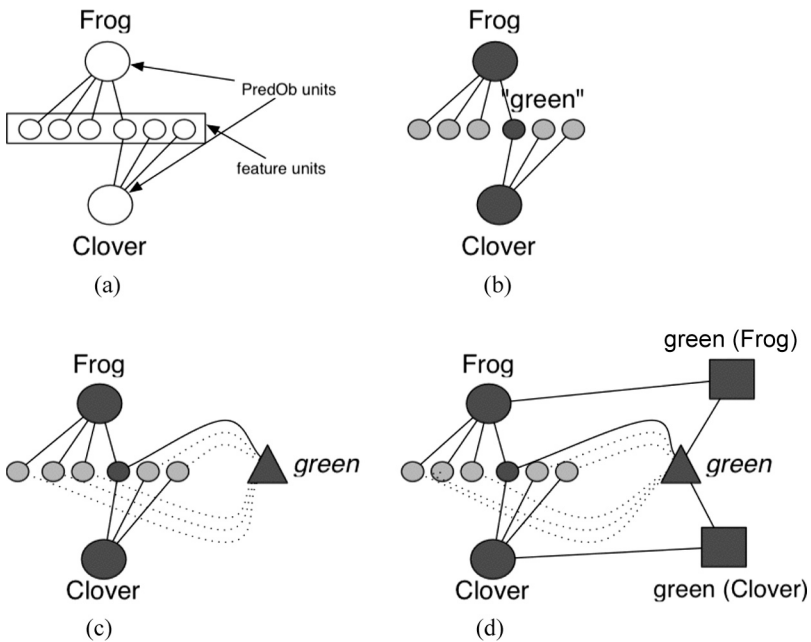


FIGURE 5 Learning in DORA. (a) A frog and a clover are compared. (b) Semantics shared by both the frog and the clover objects become more active than any unshared semantics. (c) A new unit is recruited and learns connections to the active semantics in proportion to their activation (the new unit is depicted by a triangle simply to increase the clarity of the figure). (d) The new unit is bound to the frog and to the clover via RB units. Solid lines = strong connections, dashed lines = weaker connections, darker gray = more active units, lighter gray = less active units.



(e.g., “dog,” “fierce,” etc.) and the role *big* as another PO connected to the features that define “big” bound by a RB unit that represents the binding of Fido to the property *big*.

When DORA compares two objects (e.g., a frog and a clover in Figure 5a), the PO units representing those objects become active and pass activation to their respective features. Any features shared by both objects receive approximately twice as much input and, therefore, become approximately twice as active as unshared features (here the feature “green” is shared by the frog and the clover and therefore becomes twice as active as the other features (see Figure 5b). DORA uses a self-supervised learning algorithm to learn an explicit representation of the featural overlap via a form of intersection discovery (see, e.g., Doumas et al., 2008). When two objects are compared, an “empty” PO unit is recruited and learns connections to active feature units in proportion to their activation by simple Hebbian learning (Figure 5c). Consequently, the new PO unit becomes most strongly connected to the features shared by the two compared objects because these features are the most active. That is, the new unit is an explicit representation of the shared features of the compared objects. So, following the example given in Figure 4, when a frog and a clover are compared, they activate their constituent features. Any features shared by both the frog and the clover become more active than any unshared features (in the example, the feature “green” becomes most active). When a new PO is recruited and learns connections to the active feature units most, the new PO becomes most strongly connected to the most active feature units (in Figure 5c the new PO becomes most strongly connected to the feature “green”). Finally, DORA recruits an empty RB unit in the recipient that learns a connection to active PO units (here the clover and the new representation of the property *green*; Figure 5d). The new RB unit represents the binding of the newly learned property *green* to the clover. In other words, DORA has learned an explicit representation of the greenness of the frog and the clover and has represented green as a separate and explicit feature of clover object (e.g., *green* [clover]).

In the above example, DORA learned an explicit representation of the property *green*. In the example, we used a single feature unit to code the feature “green” in order to make the example easier to follow. However, for DORA’s operation it is not important what each feature unit codes. The important aspect of DORA’s learning algorithm is that it isolates and forms explicit representations of *any* features shared by compared representations, whatever those features may be. It makes no difference to DORA whether “green” or any other feature is coded by a single feature unit or by a set of units. By comparing green things, DORA will isolate and explicitly represent the features that are invariant in green things (i.e., the features of

green however they are coded) and discard other features. For the simulation that follows, we use single feature units to code for specific features in order to make the simulations easier to follow (e.g., “green” or “blue” are coded by single feature units).

Comparison in DORA is accomplished via a mapping algorithm adopted from LISA (see Doumas et al., 2008; Hummel & Holyoak, 1997, 2003). Propositions in DORA are divided into two sets: the driver set (which controls the flow of activation in the network) and the recipient set (which responds to the pattern of activation imposed by the driver set). During mapping, one at a time POs in the driver become active and pass activation to their features. POs in the recipient compete (via lateral inhibition) to respond to the pattern of activation generated by the active driver PO. Any POs in the recipient that are connected to the active feature units will become active. The PO in the recipient that shares the most features with the active driver PO will, necessarily, become the most active. PO units that are active at the same time in both the driver and the recipient develop excitatory *mapping connections* to one another. Mapping connections represent analogical correspondences in DORA.

### *Simulation*

We simulated Study 1 as follows. Each exemplar, target, and distracter object was connected to 20 features. Although each object could, in theory, present an infinite number of features, we chose 20 as an arbitrary, but numerous, set of noticeable possibilities. That is, the small yellow shoe could be represented as “small,” “yellow,” “oblong,” “made of canvas,” “soft”, and so on. One of these features described the object’s value on the category defining dimension (i.e., the object’s color). The distracter object was created by randomly sampling from the set of features. We do not claim that this is the “right” representation of object features (e.g., that a single feature codes the color “red” or the property “oblong”). We label the features only to make the model’s performance easier to interpret; the labels themselves mean nothing to DORA. We claim only that objects have detectable properties and that these properties are separable (e.g., the same features—whatever they may be—that represent the concept “blue” become active in response to a blue object, regardless of the other features of that object like the fact that it is round).

### *Featural Similarity*

The amount of featural similarity between the objects and the exemplar was defined as the percentage of shared semantic features. There were two groups of featural similarity: one corresponding to the whole-object target

condition, and the other to the various attribute matches. In the whole-object target condition, the target object shared 80% of its semantic features with the exemplar object. That is, the two objects were almost identical (exactly as the whole-object matches, which were the same objects but different sizes). For the attribute matches, the objects were not identical but had some potential features in common. In order to simulate this condition, we matched the target and exemplar objects on 70%, 50%, 20%, or 10% of their features (selected at random). Thus the attribute matches reflected different degrees of similarity.

### *Training*

For each training trial, we placed a representation of the exemplar object into the driver and a representation of the two choice items—the target and distracter objects—in the recipient. DORA then attempted to map the exemplar object to the target and distracter objects. During mapping, the driver object became active and passed activation to its features and the target and distracter items in the recipient competed to become active. The object that most strongly mapped to the exemplar object was chosen as the winner. If the exemplar object mapped to neither object, DORA randomly selected one of the objects (i.e., it guessed). After mapping, DORA learned a representation of the featural overlap between the exemplar object and the chosen item via its intersection discovery routine. That is, DORA learned an explicit representation of what the exemplar object and chosen object had in common. If DORA selected the correct choice (i.e., chose the target item during mapping), DORA moved on to the next training trial. Alternately, if DORA chose the distracter item during mapping, DORA was required to map the exemplar object to the target object and learn the semantic overlap of these two objects via intersection discovery. DORA then built a new representation from the selected incorrect mapping and the required correct mapping via intersection discovery. This ensured that DORA not only was affected by the feedback it received when it made the incorrect choice during training but also simulated the effects of previous choices (including incorrect choices) on future performance.

On each trial (excluding the first), DORA placed a representation learned during a prior trial into the driver along with the representation of the exemplar item. Fifty percent of the time, this was the representation learned during the previous trial. The other half of the time, this was the representation learned during the last trial in which the current label was present. For example, if DORA just finished a trial where it was attempting to match two red items, DORA would compare this representation to the

representation it had learned the last time it attempted to match red items. This corresponded to the idea that children are more likely to recall more recent instances and instances in which shared labels have been presented. We chose to set correct recall at 50% because a number of studies (see Dempster & Corkill, 1999, for a review) have reported that children suffer from substantial forgetting if they are presented with new material during a retention interval. Further, because there is a large degree of variation between different studies of memory interference (as well as between children within the studies), it is difficult to determine what the “correct” forgetting percentage would be. Selecting 50% as a percentage of correct recall is in some ways artificial. However, it is well within the limits of previously reported studies of memory interference.

*The three order conditions.* We trained DORA using three training conditions analogous to the three training regimens in Study 1. Just like the participants in Study 1, DORA received four attribute match trials and two whole-object match trials per color. The order in which DORA received these trials directly mirrored the orders from Study 1. In the OM condition, DORA completed three blocks of training. The blocks were organized such that DORA learned one category per block, and in the initial trials, the target and exemplar object shared a high degree of overall featural similarity. That is, the initial trials in each block simulated the whole-object condition.

In the OD condition, DORA received the trials such that in the initial trials the exemplar object and correct target item shared the most overall similarity, but training trials from all three categories were interspersed (i.e., not blocked). In the RM training condition, categories were blocked, but trials were presented randomly within each block. Thus, in the initial trials the exemplar item did not necessarily share any more overall similarity with the target item than with the distracter item. We ran 100 simulations for each training order.

### *Test Trials*

In the test trials, DORA was presented with novel target and distracter items, each of which contained 20 features. One of these features described the object’s value on the category defining dimension. That is, if color was the defining dimension, then each object was attached to a feature describing its color (e.g., “red,” “blue,” “green”). In addition, each object was attached to 19 additional features that were selected from a pool of 100 features. As before, DORA then mapped the exemplar object to the target and distracter objects. The object that most strongly mapped to the exemplar

object was chosen as the winner. If the exemplar object mapped to neither object, DORA randomly selected one of the objects.

### *Results and Discussion*

The results of the test trials from the simulation are presented in Figure 6. Just like the children in Study 1, DORA performed most accurately on test trials when it received the OM training, and DORA was at chance when it received the OD and RM training. Importantly, DORA performs similarly regardless of which dimension defines the category. That is, the simulation results reported here suggest that the experimental results of Study 1 are not an artifact of the particular domain of color or a particular stage in development. Rather, these findings suggest that additional studies that examined other kinds of terms might find similar results, but that remains to be shown empirically.

It is helpful to understand exactly why DORA succeeds with the OM training and fails with both OD and RM training. In the OM trials, overall similarity constrains the early matches. That is, DORA (like the children in Study 1) can match the exemplar object to the target object instead of the distracter because the exemplar is simply more like the target than it is like the distracter. Because perceived similarity depends on the particular frame

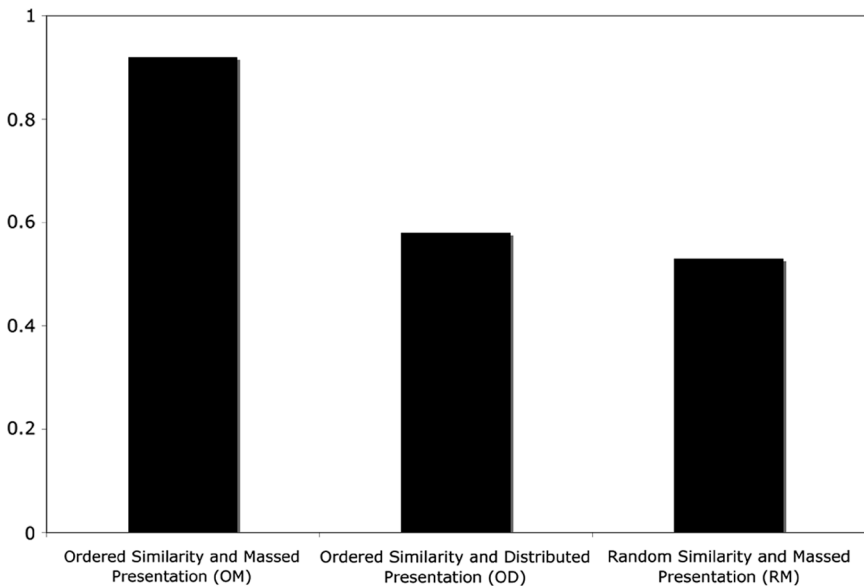


FIGURE 6 DORA's performance on the training trials in the three conditions.

of reference (Luce, 1959; Osler & Scholnick, 1968; Tversky, 1977), characteristics of both the target and the distracter objects lead to the selection of the color match. When two objects are highly similar to each other but also highly distinctive from the surrounding context, children are most likely to compare them (Paik & Mix, 2006). Accordingly, because the whole-object trials contain both high overall similarity to the color match and highly dissimilarity to the distracter object, DORA makes fewer errors on these early trials. Thus, initially the feature of interest (e.g., “redness”) comes along for the ride, so to speak, and matches are made based on holistic representations of the objects. However, over consecutive trials DORA refines its representation of the category, and over successive comparisons the extraneous features are washed out and the feature (or features) that remains consistent across all matches comes to define the category DORA is learning.

In the OD training condition, DORA (like the children in Study 1) does well on early matches again because the objects that comprise the correct matches are very similar to one another. However, in the OD training, DORA learns all categories during the same training block rather than one at a time like in the OM and RM conditions. As a result, DORA is faced with two tasks on each trial. First, DORA must select the correct item and inhibit the distracter (which may possess features that led to a correct selection in recent past trials). Second, DORA must match the current trial to the category it is currently learning—that is, it must understand that the current trial is about the color “red” rather than the color “green”—and use this information to refine its representation of “red” rather than of “green.” In the OM and RM conditions, this aspect of the task is taken care of by the order of trials. Each successive trial within a block is an instance of the same category. In the OD condition, this is not the case. Within consecutive trials, the items match for different reasons. DORA must not only make the correct match within the current trial but also link the match in the current trial to matches made in previous trials in order to refine its understanding of the category.

In the RM training condition, nothing constrains early matches. DORA (like the children in Study 1) is at chance on early trials, selecting the distracter item as frequently as it selects the match item. When DORA selects the distracter item, it cannot completely inhibit this response, and thus it uses a partial memory of the distracter item to refine its category representation. The result is that DORA learns more slowly on the trials in which it makes an error than with the trials in which it is correct. With more trials, DORA will eventually learn the category. However, as in Study 1, there is a finite number of training trials, and so it is difficult for DORA to learn to attend to the relevant feature for the category when it struggles with the early training trials.

## GENERAL DISCUSSION

In both the experimental study and the network simulation, the three conditions consisted of the exact same procedures, stimuli, and number of training trials. The *order* in which the trials were presented, however, differed between the three conditions and appeared to matter for the learner's performance. These results suggest that in learning to categorize, the order of presentation may matter as much as the content of the instances. That is, how a learner forms a specific category is determined in part by the timing in which specific category members are experienced.

Why is it that the way the task was structured in the OM condition—beginning by learning labels that apply to holistically similar things and mastering one color category before learning a second color category—resulted in rapid category learning? Learning words by viewing multiple examples requires connecting the current input with memories of previously experienced instances. There are four components of the OM condition that may aid children in inducing and learning the category.

First, the ordered similarity component may increase the ease of retrieving relevant information by connecting instances in memory. Past research has strongly indicated that learners are more likely to discover relational and abstract commonalities when given the opportunity to compare (Gentner & Namy, 1999; Namy & Gentner, 2002). Gentner (see Gentner 1983, 1989; Kotovsky & Gentner, 1996) has argued that comparison is most likely to occur when objects share high levels of surface similarity. Instances that are holistically similar are the instances that are more likely to be connected in memory. For example, if the child is presented with a red car followed by another red car and then followed by a red bus (three items that share many commonalities including color), the child is likely to group these objects together in memory, which is a first step in discovering the property of red. If instead a child is presented with a red car followed by a red sock and a red strawberry (three items that do not share many commonalities aside from color), the child is less likely to group the memory of these items together. The car may be grouped with memories of other cars (yellow cars and black cars) and the strawberry may be grouped with memories of other tasty foods (such as chocolate and bananas), but the red car and the red strawberry are less likely to be grouped in memory and, as a result, the property of red is less likely to be discovered.

Second, the massed component may help children to induce the meaning of the category because memory from one instance to the next instance is likely to be most accurate when instances are presented close together in time. Experiencing sequential examples of the same color

(e.g., red, followed by red, followed by red) ensures that particular instances of the same color category are in memory from one trial to the next, and thus the memory of past instances can be more readily retrieved and compared to subsequent examples (see Smith, Thelen, Titzer, & McLin, 1999; Spencer, Smith, & Thelen, 2001). In the memory literature, it is well established that learners are most likely to remember information that is presented through distributed, rather than massed, presentations (see Cepeda et al., 2006; Dempster, 1996, for a review). Distributed presentations may aid in remembering precise information over long time periods. However, because success in the present task requires induction of an unfamiliar category rather than straight recall or recollection of a specific instance, proximity between instances may be advantageous for making the type of comparisons that can lead to category discovery, although less optimal for remembering specific features of a single instance over a long time period.

Third, in the massed presentations, the possibility of interference from a just experienced member of a different category should be less than when instances are presented in a massed fashion. DORA, for example, compares a portion of all successive trials. When the categories are presented one at a time, these are typically appropriate comparisons (e.g., finding the similarity between a red object and another red object). However, when categories are intermixed throughout training, these are typically inappropriate comparisons (e.g., finding the similarity between a red object and a yellow object). That is, in the massed presentations, children do not need to inhibit the memory of just experienced trials.

Finally, the OM ordering may aid in psychologically reducing the pool of possible category-defining features. That is, the order of presentation may matter because as the features of objects get connected to the word "red" through multiple presentations of similar objects, over time the shared features become strongly connected to the word "red." The child's task then becomes not one of discovering the relevant feature from a large set of possible features, but one of discarding the irrelevant features from the small set of possible features shared across multiple instances. Instance similarity and massed presentations combined should together aid in reducing competing features. Indeed, the results of the training study and simulation presented here show that ordered similarity or massed presentations alone do not sufficiently improve categorization.

However, although the pattern of results in these studies suggest that children will most effectively learn new categories given a pattern of presentation similar to the OM condition, children's natural learning experiences are likely seldom ordered or massed. Yet children are facile word learners, learning on average 10,000 words by age 5 (Carey, 1978). This demonstrates



that learning categories through ordered similarity and massed training is not a prerequisite for learning. Training of other sorts will ultimately lead children to align category instances in memory, but children may require many more instances before they get to the same endpoint.

Moreover, the advantage of the type of ordered presentation found in the OM trials will arise only when the category features are ones with which the learner is unfamiliar—those that are yet to be discovered. The performance of a child who already knows many color categories (e.g., blue, pink, orange) when presented with a new color category (e.g., magenta) should be indistinguishable in any of the three training conditions. For example, Deák and Narasimham (2003) found no effect of massing or spacing with 3-year-olds when children were presented with new instances of well-known categories (e.g., object categories). The potency of the OM condition is that it leads to strengthening memory and attention to the features that matter and a weakening of memory and attention to competing irrelevant features. In the absence of ordered examples, similar effects could be accomplished by simply experiencing more examples. Furthermore, in memory studies, subjects typically learn more rapidly when the material is massed, but they remember the material for shorter periods (Schmidt & Bjork, 1992). This could be a factor in the present study because subjects were tested shortly after training, and thus, it is unclear how long subjects were able to retain what they learned during training.

So why is less sometimes more in the case of categorization tasks? There are two aspects involved in category learning. First, learners must link instances that share categorical membership, and second, learners must define category boundaries by identifying the features that matter and excluding the irrelevant features (see Rosch, 1977). Experiences that emphasize only one of the aspects but not the other are likely to either inhibit category formation or result in category overextensions and underextensions. Additionally, the results of these studies suggest that category learning may be more successful when learners do not attempt to master both category learning aspects simultaneously. Instead, category learning may be most successful when these aspects are dealt with one at a time. First, instances that form a category are grouped through strong holistic similarity. In this way, starting with less variation in the initial experienced instances of a category may contribute to a stronger, more cohesive categorical grouping. Second, category boundaries are refined through successive comparisons of category instances. The present results suggest that learners are more successful when fewer categories are simultaneously presented—that is, when instances are presented in close proximity without interference from other categories. In sum, the idea that starting with less variation between experienced exemplars more readily leads to learning appears to

go beyond the language acquisition literature and may describe a more general-purpose approach to learning complex problems.

## REFERENCES

- Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 18*, 513–562.
- Carey, S. (1978). The child as word learner. In J. Bresnan, G. Miller, & M. Halle (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.
- Casasola, M. (2005). When less is more: How infants learn to form an abstract categorical representation of support. *Child Development, 76*, 279–290.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354–380.
- Deák, G. O., & Narasimham, G. (2003). Is perseveration caused by inhibition failure? Evidence from preschool children's inferences about word meanings. *Journal of Experimental Child Psychology, 86*, 194–222.
- Deák, G. O., Ray, S. D., & Pick, A. D. (2002). Matching and naming objects by shape or function: Age and context effects in preschool children. *Developmental Psychology, 38*, 503–518.
- Dempster, F. (1996). Distributing and managing the conditions of encoding and practice. In E. Bjork & R. Bjork (Eds.), *Handbook of perception and cognition* (2nd ed., pp. 317–344). New York: Academic Press.
- Dempster, F., & Corkill, A. J. (1999). Interference and inhibition in cognition and behavior: Unifying themes for educational psychology. *Educational Psychology Review, 11*, 1–88.
- Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*, 1–43.
- Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. New York: Dover.
- Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition, 12*, 20–30.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71–99.
- Fanselow, M. S., & Tighe, T. J. (1988). Contextual conditioning with massed versus distributed unconditional stimuli in the absence of explicit conditional stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 187–199.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.
- Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59*, 47–59.
- Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). London: Cambridge University Press.
- Gentner, D., & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development, 14*, 487–513.
- Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development*. Cambridge, UK: Cambridge University Press.
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427–466.

- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*, 220–264.
- Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 829–846.
- Kersten, A. W., & Earles, J. L. (2001). Less really is more for adults learning a miniature artificial language. *Journal of Memory and Language*, *44*, 250–273.
- Kersten, A. W., & Smith, L. B. (2002). Attention to novel objects during verb learning. *Child Development*, *73*, 93–109.
- Kotovskiy, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*, 2797–2822.
- Luce, R. D. (1959). *Individual choice behavior*. New York: John Wiley.
- Medin, D. L., & Bettger, J. G. (1994). Presentation order and recognition of categorically related examples. *Psychonomic Bulletin & Review*, *1*, 250–254.
- Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sows' ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, *131*, 5–15.
- Newport, E. (1990). Maturation constraints on language learning. *Cognitive Science*, *14*, 11–28.
- Osler, S. F., & Scholnick, E. K. (1968). The effect of stimulus differentiation and inferential experience on concept attainment in disadvantaged children. *Journal of Experimental Child Psychology*, *6*, 658–666.
- Paik, J. H., & Mix, K. S. (2006). Preschoolers' use of surface similarity in object comparisons: Taking context into account. *Journal of Experimental Child Psychology*, *95*, 194–214.
- Pazzani, M. (1991). The influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology—Learning, Memory & Cognition*, *17*(3), 416–443.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353–363.
- Rice, N. (1980). *Cognition to language*. Baltimore, MD: University Park Press.
- Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*, 67–109.
- Rosch, E. (1977). Human categorization. In N. Warren (Ed.), *Advances in cross-cultural psychology* (Vol. 1). London: Academic Press.
- Ross, B. H. (1996). Category representations and the effects of interacting with instances. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1249–1265.
- Ross, B. H. (1997). The use of categories affects classification. *Journal of Memory and Language*, *37*, 240–267.
- Sandhofer, C. M., & Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology*, *35*, 668–679.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*, 207–217.
- Smith, L. B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of acting: The task dynamics of the A-not-B error. *Psychological Review*, *106*, 235–260.
- Spencer, J. P., Smith, L. B., & Thelen, E. (2001). Tests of a dynamic systems account of the A-not-B error: The influence of prior experience on the spatial memory abilities of 2-year-olds. *Child Development*, *72*, 1327–1346.
- Toppino, T. C. (1991). The spacing effect in young children's free recall: Support for automatic-process explanations. *Memory & Cognition*, *19*, 159–167.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

- Underwood B. J. (1961). Ten years of massed practice on distributed practice. *Psychological Review*, *68*, 229–247.
- Vander Linde, E., Morrongiello, B. A., & Rovee-Collier, C. (1985). Determinants of retention in 8-week-old infants. *Developmental Psychology*, *21*, 601–613.
- Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Wattenmaker, W. D. (1993). Incidental concept learning, feature frequency, and correlated properties. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 203–222.
- Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology*, *18*, 158–194.
- Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language*, *39*, 124–148.

APPENDIX  
Objects Used in the Training Trials

<i>Exemplar</i>	<i>Color match</i>	<i>Distracter</i>
Whole-object similarity matches		
Large red cup	Red cup	Green cowboy
Large red cup	Red cup	Yellow hat
Large green dino	Green dino	Red hammer
Large green dino	Green dino	Yellow cone
Large yellow shoe	Yellow shoe	Red flower
Large yellow shoe	Yellow shoe	Green horse
Attribute matches		
Large red cup	Red apple	Yellow wrench
Large red cup	Red apple	Green rocking horse
Large green dino	Green heart	Red tongue depressor
Large green dino	Green heart	Yellow cheese
Large yellow shoe	Yellow ring	Green plate
Large yellow shoe	Yellow ring	Red train
Large red cup	Red tube	Yellow tube
Large red cup	Red tube	Green tube
Large green dino	Green basket	Red basket
Large green dino	Green basket	Yellow basket
Large yellow shoe	Yellow table	Green table
Large yellow shoe	Yellow table	Red table