Why Children Learn Color and Size Words So Differently: Evidence From Adults' Learning of Artificial Terms

Catherine M. Sandhofer and Linda B. Smith
Indiana University

An adult simulation study examined why children's learning of color and size terms follow different developmental patterns, one in which word comprehension precedes success in nonlinguistic matching tasks versus one in which matching precedes word comprehension. In 4 experiments, adults learned artificial labels for values on novel dimensions. Training mimicked that characteristic for children learning either color words or size words. The results suggest that the learning trajectories arise from the different frames in which different dimensions are trained: Using a comparison (size-like) training regimen helps learners pick out the relevant dimension, and using a categorization (color-like) training regimen helps the learner correctly comprehend and produce dimension terms. The results indicate that the training regimen, not the meanings of the terms or the specific dimensions, determines the pattern of learning.

The idea that the language we hear determines the thoughts we have is both profound and, in its usual form, highly contentious. Are there conceptual distinctions without corresponding linguistic distinctions? Are human concepts universal? In this article, we consider a much less contentious but perhaps equally profound idea—that the language we hear subtly influences the thoughts we have. Specifically, different ways of talking about concepts may engage distinct learning processes that in turn lead to different learning. In this way, language may help to shape both processes and the outcome of learning. This idea is consistent with findings from studies of adult category learning that suggest that how information is structured ultimately determines what is learned (Lamberts, 1994; Posner & Keele, 1968; Ross, 1997; Schyns, Goldstone, & Thibault, 1998; Schyns & Rodet, 1997; Tversky, 1977; Waldmann & Holyoak, 1992; Ward & Becker, 1992; Wattenmaker, Dewey, Murphy, & Medin, 1986; Yamauchi & Markman, 1998). We propose that these findings about the role of task structure in adult learning are related to the role of language in cognitive development. That is, cognitive development may be influenced by the task structures that emerge in different linguistic environments.

Our starting point is recent evidence, indicating that children learn color and size terms in very different ways (Sandhofer & Smith, 1999). We asked the following: Why are the developmental patterns different in the two cases? Our results suggest that the cause is neither specific to the developmentally young nor to these

---

particular dimensions. Instead, the different developmental patterns seem to arise from the different ways we talk about colors and sizes.

Learning Dimensional Adjectives

Dimensional adjectives are the class of words that refer to the perceptible properties of individual objects—words like *little*, *blue*, *wet*, and *soft*. As a class, dimensional adjectives are learned slowly and errorfully by young children (Carey, 1982; Gasser & Smith, 1998; Nelson, 1973; L. B. Smith & Sera, 1992).

Children's slow acquisition of color words has been a puzzle for some time. Early in development, children seem almost incapable of learning color terms. For instance, in one study Rice (1980) reported that children required an average of 1,080 trials to learn the three color terms "red," "green," and "yellow" (see also Andrick & Tager-Flusberg, 1986). Children's early difficulties with color words seem so profound, that Charles Darwin actually speculated that children are initially color blind (Darwin, 1877). Other work has shown that children in the early stages of learning often use individual color words such as "red" or "blue" but without apparently mapping them to the appropriate color categories. That is, although young children answer the question "What color is it?" with a color term, the color terms they provide seem to be randomly chosen and unrelated to the property in question (Backscheider & Shatz, 1993; Bartlett, 1978; Cruse, 1977; Istomina, 1963; Landau & Gleitman, 1985; Sandhofer & Smith, 1999). This could mean that children first understand that color words refer to color without knowing how color words map to specific color categories (Carey, 1982). Alternatively, at this early stage, children might only know that words such as "red" or "blue" are appropriate answers to "What color is this?" without knowing why. This second idea—that children answer questions about color with color words because of mere linguistic associations and not conceptual knowledge—is supported by the fact that children who are blind from birth and thus have no direct experience of colors also answer these questions in the very same way (see Landau &

Gleitman, 1985; see also Shepard & Cooper, 1992, for evidence for this idea in a blind child). In sum, the developmental progress in learning color words seems to move from learning about words to learning about the properties to which those words refer. Indeed Soja (1994) suggested that early in color-word learning, children do not realize that colors are nameable properties.

Children also acquire size terms slowly, but the pattern is very different from that in the case of color. "Big" and "little" are among the first dimension terms used productively, and at a global level they are used correctly from quite early. Specifically, children use "big" to talk about big things and "little" to talk about little things, and in some experimental studies children as young as 2 years have performed perfectly in tasks that assess comprehension of the size terms "big" and "little" (Carey, 1982). However, by other measures, size words are not fully understood until around 5 years of age. Even in the late preschool years, children seem to treat size words as category terms instead of relative terms, and the understanding of size terms as comparisons along a continuum is late (Clark, 1970; L. B. Smith & Sera, 1992). The tendency to use these terms categorically also is evident in children's resistance to shifting standards. Once an object is labeled *big*, for example, children have difficulty acknowledging that it is smaller than some other objects even when the size difference is considerable (Gitterman & Johnston, 1983; Sera & Smith, 1987). Thus, in the case of size terms, children seem to first realize that sizes are nameable properties and then slowly work out the intricacies of how these terms are used relationally.

The Phenomenon

In this article, we focused on a specific difference between size- and color-word learning, reported by Sandhofer and Smith (1999). In a 6-month longitudinal study, Sandhofer and Smith followed children's acquisition of color and size terms, using three measures: (a) correct responding in a comprehension test, (b) linguistic associations between questions like "What color/size is it?" and specific property terms, and (c) matching by the dimension in a nonlinguistic task. The results of the study suggest that there is a typical order of success on the three measures. However, the characteristic order for color acquisition is different than the characteristic order for size acquisition. Figure 1 shows the mean age at which children succeeded in each of the three component measures for color and size.

In learning color words, children appear to first learn associations among color words. That is, they can answer the question "What color is this?" with a color term (albeit not necessarily a correct term). Next, children appear to succeed in comprehension tasks that ask them to pick out objects of specific named colors, for example, "Show me the red one." Finally, and often a long time into this already protracted developmental course, children successfully match objects (that differ on other properties) by color in nonlinguistic tasks. This seems paradoxical in that it suggests that selective attention to color, as evidenced by success on the matching trials, is not a prerequisite for learning color names, but rather it developmentally follows the learning of color names.

In contrast, most children first succeeded on the size task that required matching objects. That is, before they knew size words well, children selectively attended to size in a nonlinguistic task and matched objects that differed in other ways by size. Shortly after, children reliably picked the bigger of two objects when asked to select the "big" one and the smaller of two objects when asked to select the "little" one. It is only very late in the acquisition process that children make linguistic associations between the word "size" and specific size words by answering with a size term when the experimenter pointed to one object and asked (in an analogous way to the most common question children are asked about color), "What size is this?"
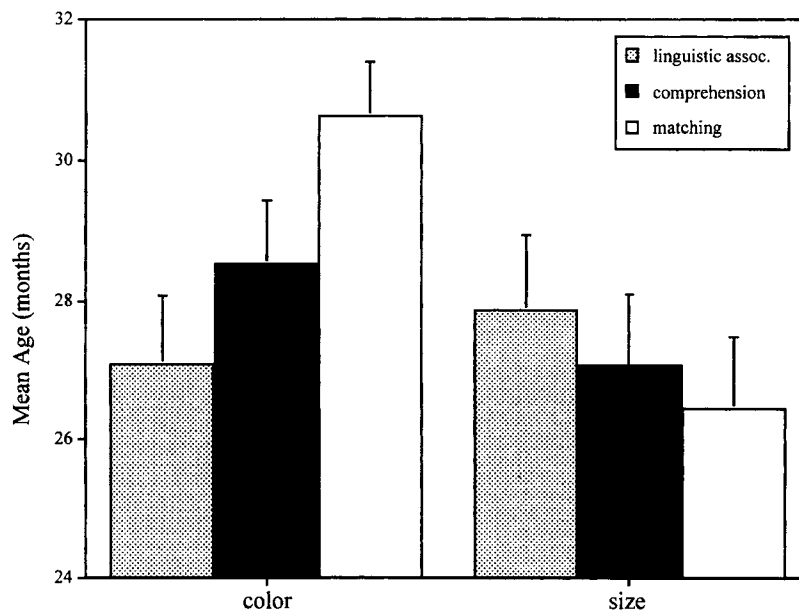

Figure 1. Mean age of acquisition in the three measures of color and size knowledge (data are from Sandhofer & Smith, 1999). assoc. = association.

The present study specifically seeks to explain the contrast between performance on the word-comprehension task and the nonlinguistic matching task. The literature on children's acquisition of dimensional terms is replete with comparisons of children's performances in language comprehension tasks and nonlinguistic matching tasks. The theoretical questions behind these comparisons concern the independence or interactions between language learning and perceptual development. Is children's learning of dimension terms simply a matter of mapping words to already available percepts or does language learning help construct dimensional knowledge? Findings that children can match by a dimension in a nonlinguistic task prior to being able to label that dimension or its properties have been interpreted as evidence that children possess a conceptual representation of the dimension, but they have difficulties mapping words to this representation (Markman, 1989; Soja, 1994; Waxman & Gelman, 1986). In contrast, findings that children can label a dimension or its properties prior to being able to match by that dimension have been interpreted as evidence that children are unable to selectively attend to the dimension without verbal mediation (Kendler & Kendler, 1961) or even that the process of learning the dimension labels constructs the dimensional concepts (L. B. Smith, Gasser, & Sandhofer, 1997). By these interpretations, the different developmental trajectories that characterize color and size learning—words before matching or matching before words—implicate very different roles for language. In the case of color, word learning seems to drive abstraction of the dimension; in the case of size, words seem to map onto an existing conceptual structure. The results we present from the adult simulation studies challenge these interpretations.

Possible Sources of Difference: Language and Task Structure

There are many differences between color and size that could lead to the different learning trajectories: (a) differences in the meanings of the terms (e.g., color terms refer to categories [i.e., a red object is red in all contexts], whereas size terms refer to relations [i.e., a big object is big in some comparison contexts and little in others]; Clark, 1970; Gleitman & Wanner, 1988; Martin, 1969; Vendler, 1968); (b) differences in the ways the dimensions are perceived by the sensory system (Adams, 1989; Brown & Teller, 1989; Heider, 1971; Slater & Johnson, 1998); (c) differences in the numbers of common color and size terms (Park, Tsukagoshi, & Landau, 1985); and (d) differences in the way the dimensions are commonly talked about (taught) by caregivers. In this article, we concentrated on this last difference, on the effect of the different ways that color and size are commonly presented in speech to children.

Different frames for talking about color and size are pervasive in the input to children (Sandhofer, 2001). Table 1 summarizes these differences in the ways colors and sizes are talked about. Color words are taught by presenting a single object and asking "What color is this?" This type of input emphasizes color words as category labels. It also teaches children that questions about color demand an answer from a circumscribed set. Notice that these queries essentially constitute a production task based on a single object. In contrast, size words are taught by presenting multiple objects and asking for example, "Where's the big one?" This type

Table 1

Three Differences in Task Structure and Meaning in Color and Size Learning

	Dimension	
Color	Size	
Categorical	Relational	
Production	Comprehension	
Single object	Comparison among multiple objects	

of input encourages children to compare items in a set and select the one that best fits the provided label. We suspect that no one ever points to a single object and asks the child (as they do in the case of color) "What size is this?" Instead, parental queries about size constitute a comprehension task that emphasizes comparisons among multiple objects.

Could these differences in the ways colors and sizes are talked about and taught to young children create the different developmental trajectories? They are certainly consistent with the early acquisition of linguistic associations between the question "What color is this?" and the set of color words, relative to that between the question "What size is this?" and the set of size words. Could these differences also lead to color-word knowledge that precedes attention to similarities and differences in color but attention to similarities and differences in size that precede size-word learning?

In the reported experiments, we sought to disentangle the potentially independent contributions of task structure (i.e., type of training and number of objects referred to) and meaning (i.e., categorical vs. relational). If either the structure of the learning task or the meanings of the terms are critical—not the developmental level of the learner and not the particular sensory and perceptual dimension—then we should be able to experimentally create the two different learning trajectories in adults, by teaching them in different ways: color-like meanings in a color-like production task or size-like meanings in a size-like comprehension task.

Current Experimental Issues

The present experiments were designed to test the hypothesis that different task structures are instrumental in creating either a learning pattern in which selective attention to the dimension precedes lexical learning or a learning pattern in which lexical learning precedes selective attention to the dimension. To test this idea, we conducted a human simulation experiment, using college undergraduates as the experimental population (see Gillette, Gleitman, Gleitman, & Lederer, 1999, for a discussion of using humans as a simulation device). We chose adults as our experimental population to eliminate the possibility that the learning trajectories observed in children were because of maturational or conceptual limitations (Bornstein, 1985).

We also addressed the possibility that the developmental patterns are dimension specific, that is, the learning trajectory children commonly exhibit when learning color words is a product of learning about color stimuli and the learning trajectory children exhibit when learning about size is a product of learning about size

stimuli. We did so by presenting participants with novel stimuli. In the first two experiments, we presented participants with the same novel dimensions in two different training conditions: one that simulates the production task structure and categorical meaning of color terms and one that simulates the comprehension task structure and relational meanings of size terms.

In each experiment, the adult participants were taught six novel adjectives that refer to categories or relations on two dimensions. Participants learned about two dimensions simultaneously because children commonly learn multiple sets of dimensional terms at the same time in their natural environments. For example, given a picture of a big yellow duck and a small green car, children may be asked "What color is that?", "Where's the little one?", and "Which one is soft?" all within a brief period of time. Children must learn to switch attention between relevant features to correctly learn the words, and thus we asked adults to solve this same learning problem.

In Study 1, participants learned novel words that refer to values on the integral dimensions of brightness and saturation. Integral dimensions are those that are difficult even for adults to perceive independently (Garner, 1974). Such dimensions arguably provide a better analogue to children's learning about color and size, because children generally cannot selectively attend to one dimension (L. B. Smith, 1989). In Study 2, the same procedure was used except that participants were taught names that referred to categories and relations on the separable dimensions of brightness and angle degree. Because these dimensions can be perceived independently, they should be separable for adults. In Studies 3 and 4, we examined the relative contributions of categorical versus relational meaning and production versus comprehension task structure.

Study 1

In Study 1, half of the participants were presented with a learning task designed to be analogous to color learning: On each trial, dimension terms with categorical meaning were taught in a production task in which participant attention was focused on one object. We called this training Condition C (for color). The other half were presented with a learning task designed to be analogous to size learning: Dimensional terms with relational meanings were taught in a comprehension task in which participants had to compare objects to make a choice. We called this training Condition S (for size). In both cases, participants learned words that map to values or relations on the integral dimensions of brightness (value) and saturation (chroma). Figure 2 illustrates the training procedure used in the two conditions. In Condition C, a linguistic context analogous to that used when children are taught colors ("What color is that?") was used to train the novel dimensions. On each trial, three stimulus items were presented, and the experimenter pointed to one and asked for example "What chroma is this?" The participant's task was to provide the experimental name (e.g., "dax") and feedback was provided. Moreover, each of the to-be-learned words refer to categories of fixed values on one dimension.

The training procedure in Condition S was different and designed to be analogous to children's experiences in learning size words. Thus, in Condition S, the linguistic context commonly presented in learning size words ("Where's the big one?") was used. On each trial, three stimulus items were presented, and the experimenter asked for example "Where's the dax one?" Moreover, in Condition S the to-be-learned words referred to the direction of difference on one dimension. That is, the same saturation value was not always "the rif one," because rif referred to the least saturated value of those presented.

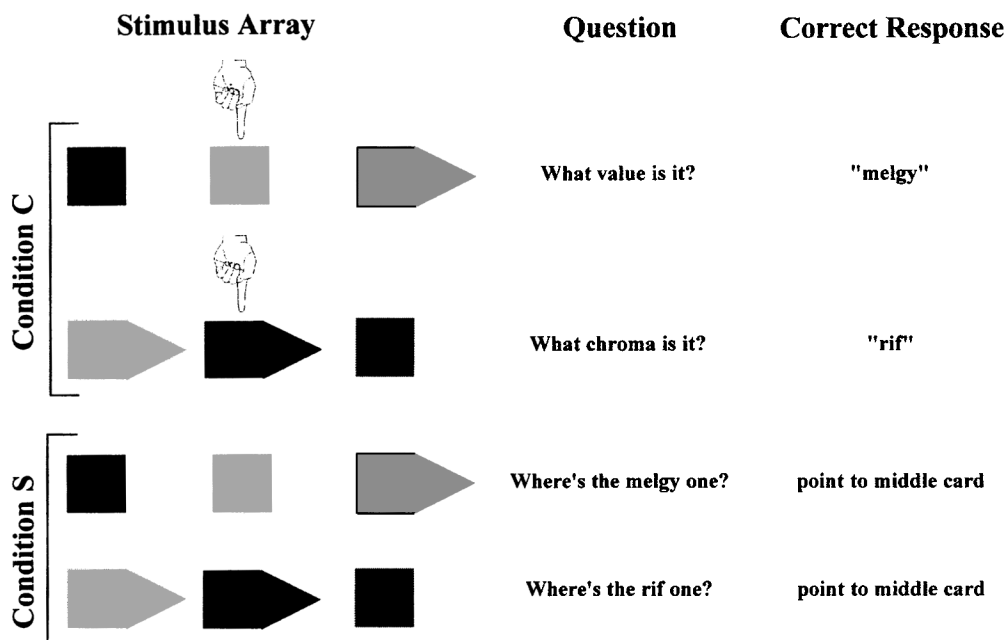| Stimulus Array	Question	Correct Response
Condition C	What value is it?	"melgy"
Condition C	What chroma is it?	"rif"
Condition S	Where's the melgy one?	point to middle card
Condition S	Where's the rif one?	point to middle card

Figure 2. Illustration of the two training conditions used in Studies 1, 2, 3, and 4. C = color; S = size.

Thus, in Experiment 1, just as in children's experiences learning color and size words, there were many training differences. Condition C used a linguistic cue ("What chroma/value is it?") to signal the relevant dimension; the task is a production task; and the terms map to categories. Condition S, in contrast, used only the six property terms themselves, no label that signals the dimension; the task is a comprehension task; and the terms refer to relations among the levels presented on the queried dimension. Thus, the results of this experiment cannot tell us which of these factors matter, but they addressed what we see as the prior question: If we teach adults dimensional terms in a color-like way, does their progress in learning mimic the developmental trend in learning color words? If we teach adults dimensional terms in a size-like way, does their progress in learning mimic the developmental trend in learning size words? More specifically, does Condition C result in the late acquisition of matching objects on the trained dimension, whereas Condition S results in the early acquisition of matching objects on the trained dimension?

Method

Participants

Twenty-eight native English-speaking undergraduates from Indiana University participated in this experiment in partial fulfillment of introductory psychology course requirements. In each condition, half of the participants were female and half were male. Each participated in four 1-hr sessions.

Materials and Design

One-by-one-inch squares of Munsell 10R hue paper and 1 × 2-in. pentagons of Munsell 10B hue paper were pasted in the center of 3 × 5-in. index cards. For both Munsell hues, the values (brightness) presented to participants consisted of all available values between 3 and 8 and the

chromas (saturation) consisted of all available chromas between 1 and 10. The stimuli thus formed a 2 × 6 × 6 matrix, with each hue combined with every available brightness level and every available saturation level.

In both conditions, the participants were required to learn words that referred to categories or relations among brightnesses and saturations. However, to make the selective attention task more difficult and more like that encountered by children, variation on two additional irrelevant dimensions, hue and shape, was added. Specifically, the chromas and values were realized as red (10R) squares or blue (10B) pentagons.

In Condition C, the three categories on each dimension were defined by fixed levels. For value (brightness), they were low brightness (values 3 and 4), medium brightness (values 5 and 6), and high brightness (values 7 and 8). For chroma (saturation), they were low saturation (chromas 1 and 2), medium saturation (chromas 4 and 6), and high saturation (chromas 8 and 10). The particular verbal tokens assigned as the label for each category (dax, wug, melgy, zup, slee, rif) were randomly assigned. In Condition C, participants were presented with three stimulus cards on each trial, but only one was pointed to and participants were asked to label the one card to which the experimenter pointed. Figure 3 (left panel) specifies the mapping of labels to stimulus categories in Condition C. Note, nine unique categories are defined by the conjunction of two category labels, one referring to the attributes on one dimension and the other referring to attributes on the other dimension. However, each value–chroma combination spans eight unique combinations of hue, saturation, and brightness values; for example, the combination "rif/wug" encompasses the eight hue–value–chroma combinations of (in Munsell notation) 10R 3,1; 10R 3,2; 10R 4,1; 10R 4,2; 10B 3,1; 10B 3,2; 10B 4,1; and 10B 4,2.

In Condition S, three relations were defined on each dimension: lowest level present, the medium level present, and highest level present. Figure 3 (right panel) specifies the mapping of labels to stimulus relations in Condition S. Three words referred to these relations on the value dimension and three referred to relations on the chroma dimension. For example, of the three cards simultaneously presented, the card with the lowest brightness was labeled "wug," the card with the medium brightness was labeled "dax," the card with the highest brightness was labeled "melgy." However,
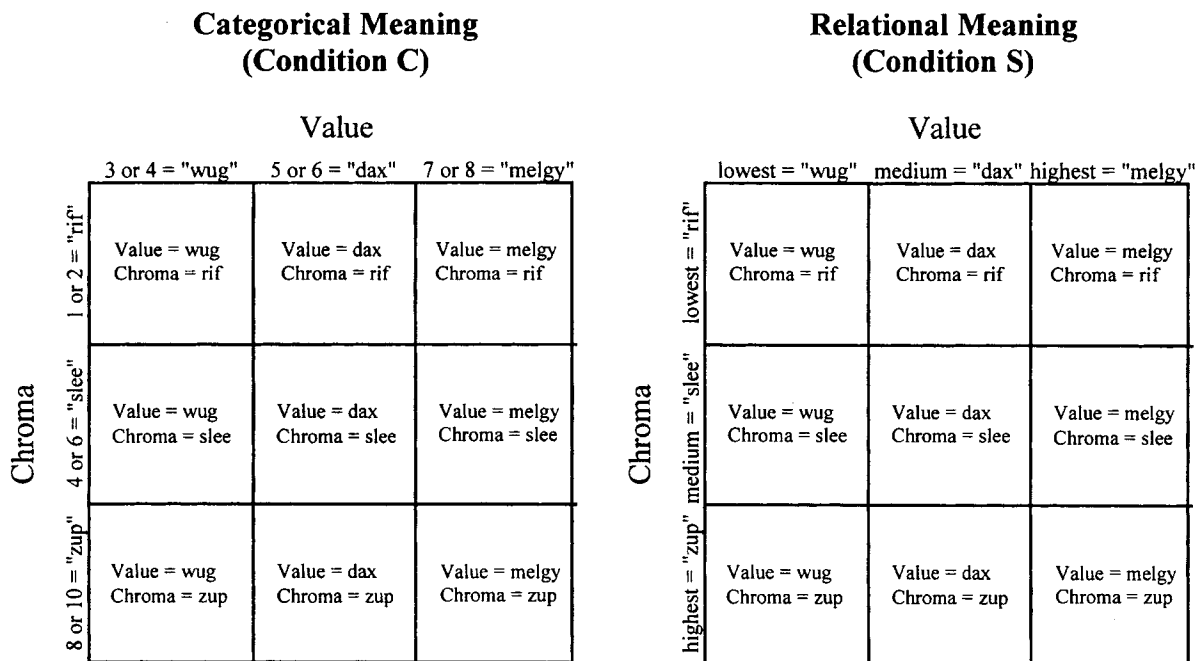
Categorical Meaning (Condition C)

Value

	3 or 4 = "wug"	5 or 6 = "dax"	7 or 8 = "melgy"
1 or 2 = "rif"	Value = wug Chroma = rif	Value = dax Chroma = rif	Value = melgy Chroma = rif
4 or 6 = "slee"	Value = wug Chroma = slee	Value = dax Chroma = slee	Value = melgy Chroma = slee
8 or 10 = "zup"	Value = wug Chroma = zup	Value = dax Chroma = zup	Value = melgy Chroma = zup

(Chroma — left vertical axis label)

Relational Meaning (Condition S)

Value

	lowest = "wug"	medium = "dax"	highest = "melgy"
lowest = "rif"	Value = wug Chroma = rif	Value = dax Chroma = rif	Value = melgy Chroma = rif
medium = "slee"	Value = wug Chroma = slee	Value = dax Chroma = slee	Value = melgy Chroma = slee
highest = "zup"	Value = wug Chroma = zup	Value = dax Chroma = zup	Value = melgy Chroma = zup

(Chroma — left vertical axis label)

Figure 3. The mappings of labels to stimulus categories in Study 1. C = color; S = size.

the specific brightness that was low, medium, or high varied across trials. Analogously, the card with the lowest saturation was labeled "rif," the card with the medium saturation was labeled "slee," and the card with the highest saturation was labeled "zup." However, the specific saturations that were low, medium, or high varied across trials. During training, the three cards that were presented on every trial were arranged so that each available value and chroma were presented equally often as the lowest, medium, and highest of the three cards. However, in some instances certain cards could not be presented in all three relational positions across trials. For example, values of 10 (the highest available brightness) were never presented as the low or medium brightness of the three cards. Participants were presented with three cards simultaneously and asked to select one as the correct answer.

In both conditions, participants were presented with 63 training trials at each session—seven trials from each of the nine possible value–chroma label combinations (wug–rif, wug–slee, wug–zup, dax–rif, dax–slee, dax–zup, melgy–rif, melgy–slee, melgy–zup). In both conditions nine cards, one randomly chosen from each of the value–chroma label combinations of Condition C were reserved for the testing portions of the experiment, and participants were not trained to name these nine unique combinations of value, chroma hue, and shape. In instances where a value–chroma combination was not available from Munsell (e.g., the combination of value 8 and chroma 10 in 10R hue was not available), a randomly selected stimulus from that value–chroma label combination was used twice. This occurred in six cases: three in the 10R hue and three in the 10B hue. The order of the test trials was randomly determined.

Procedure

Participants were trained and then tested once each day for 4 consecutive days. Each participant was tested individually by an experimenter. Participants were told that we were interested in learning how well they could learn new words when they were taught to learn words by hearing multiple examples named, as a toddler learns words. At every session, participants completed a training session followed by a testing session. Table 2 shows the sequence of events over the course of the experiment. In addition, participants completed a nonverbal matching pretest at the beginning of the first session.

Table 2
Sequence of Procedures in Experiments 1–4

Session no. and test	No. of trials
1	
Pretest	4
Training Session 1	63
Production	4
Comprehension	4
Matching	4
2	
Training Session 2	63
Production	4
Comprehension	4
Matching	4
3	
Training Session 3	63
Production	4
Comprehension	4
Matching	4
4	
Training Session 4	63
Production	4
Comprehension	4
Matching	4

Training

During each trial of Condition C, three cards were placed in front of the participant. The center card was pointed to by the experimenter and the participant was asked either "What value is this?" or "What chroma is this?" If the participant answered correctly the experimenter responded "You're right. That value is wug." If the participant answered incorrectly the experimenter responded "No. That value is wug."

During each trial of Condition S, three cards were placed in front of the participant and the participant was asked to, for example, "Show me the wug one." If the participant selected the correct card the experimenter responded "You're right. That's the wug one." If the participant selected the incorrect card the experimenter responded "No. This is the wug one," and pointed to the correct card.

Testing

One issue in designing this study is how to test and properly compare the performance of the two groups given that they were trained in very different ways, one group with production training and one group with comprehension training. One possibility is to compare the two groups' knowledge of specific words by comparing performance in a task of the type they were trained with, such that the participants trained with comprehension training would be given a comprehension word test and the participants trained with production training would be given a production word test. This seems inappropriate because the two types of tests themselves are not comparable and might yield different performance levels despite similar underlying knowledge. Accordingly, our solution was to compare performance on the same type of test for both groups. We chose to use a comprehension test because we reasoned that participants trained with both comprehension training and production training would be able to demonstrate word knowledge in a comprehension format, whereas participants in a comprehension-training group would experience difficulty with a production test. However, given the possibility that participants trained with a production-training regimen would only perform well on the production test, we also included this test, and to ensure equal testing for both groups, we included a production test both for participants in Condition C and Condition S. Given both groups' greater success on the comprehension test rather than the production test, we do not analyze the results of the production test in any great detail.

Table 2 shows the sequence in which each test was presented at each session. During each session of the four sessions of the experiment, participants completed a dimension production test, followed by a dimension comprehension test, and finally a nonverbal matching test. Participants were not given feedback during the testing sessions.

Production test. Participants completed four production test trials. The production test was identical to the training trials in Condition C. Three cards were placed in front of the participant. The center card was pointed to by the experimenter, and the participant was asked either "What value is this?" or "What chroma is this?" The production test trials were identical for both Condition C and Condition S; however, the way in which the correct answer was determined was dependent on the training condition— categorically scored for Condition C and relationally scored for Condition S. Whether an answer was correct or incorrect was used for scoring purposes only—participants were not given feedback on the test. The adjectives selected to be queried in a session and the stimuli presented were randomly determined for each session. Two of the stimuli were untrained, that is, they had not been presented in training, and two were training stimuli, that is, they had been presented in training.

Comprehension test. Participants completed four comprehension test trials. The comprehension test was identical to the Condition S training trials. Three cards were placed in front of the participant, and the participant was asked to, for example, "Show me the wug one." The comprehension test trials were identical for both Condition C and Condition S. The adjectives queried in a session and stimuli presented were randomly

determined. Participants were not provided with feedback on the test. Two of the stimuli were untrained, that is, they had not been presented in training, and two were training stimuli, that is, they had been presented in training.

Nonverbal matching test. The logic behind the nonverbal matching task was to present participants with a task that would require them to selectively attend to the relevant dimension but would not require dimensional language. In the matching trials the experimenter first laid out one row of three stimuli cards before the participant. The cards varied in shape, hue, brightness, and saturation. The experimenter then laid out a second row of three stimuli cards that also varied in shape, hue, brightness, and saturation and indicated that the center card in the first row matched the center card in the second row. Table 3 shows one example of the matching test for the two conditions. For the participants in Condition C, the two exemplars matched by being either the identical value or the identical chroma. For participants in Condition S the two exemplars matched by being, for example, both the darkest in their respective sets of three. The participant was then given three choice cards and asked to find the one that matched these two center exemplars in the same way. In Condition C, one of the choice cards matched the exemplars by having either the identical value or identical chroma. In Condition S, one of the choice cards matched the exemplars by being, for example, also the darkest of the three choice cards. At each test session, participants completed four of these matching test trials. Two of the stimuli were untrained, that is, they had not been presented in training, and two were training stimuli, that is, they had been presented in training.

Pretest. The pretest given at the beginning of the first session was identical to the nonverbal matching test.

Results

Pretest

The pretest results showed that participants in the two training conditions could not match brightnesses and saturations prior to training. More specifically, when presented with two exemplars that matched in brightness or saturation, participants in Condition C were, prior to training, unable to select a third object of the same value, choosing the matching object on average 43% of the time, which did not differ from chance. Similarly, when presented with two relationally specified examples (e.g., each the lightest in their respective sets of three), participants in Condition S were unable to select the target value that stood in the same relation within its set, choosing the relational match 36% of the time, which also did not differ from chance. These chance level performances in the matching pretest tell us that the categorical and relational matches on these two dimensions are not immediately obvious to participants.

Training

Before turning to the main question, we examined participants' choices during the training trials. The mean proportion correct for the four training trials is listed in the Appendix. We asked how well participants learned the property terms in the two training conditions by examining the proportions of correct responses. A 2 (training condition) × 2 (dimension: value–chroma) × 4 (session) analysis of variance (ANOVA) yielded a main effect of dimension, $F(1, 26) = 20.03$, $p < .01$, a main effect of session, $F(3, 24) = 23.539$, $p < .01$, but no main effect of condition. The ANOVA also yielded a significant interaction of Session × Dimension, $F(3, 24) = 6.778$, $p < .01$. Overall, as illustrated in Figure 4, participants tended to learn values before chromas but learned the dimension words approximately equally well in both conditions.

In Condition C, participants also had the opportunity to learn associations between the names for the dimensions—value and chroma—and the property terms. The evidence suggests they did so readily. Just as young children rapidly learn to answer the question "What color is it?" with a color word (even when they cannot map names to colors) adults also rapidly learned to answer the question "What chroma is it?" with one of the chroma terms and the question "What value is it?" with one of the value terms (even though they often selected the wrong term early in training). Indeed, during the first training session, adults answered chroma questions by supplying chroma property labels and answered value questions by providing value property labels on 82% of training trials. By the end of Session 1, participants in Condition C knew which three of the six terms were chroma terms and which were value terms.

Testing

We first report performance on the tests at the end of learning. Because participants in Condition S had never been exposed to the words *value* and *chroma* in the production question "What value/chroma is this?", participants in Condition S performed considerably worse on the production test than participants in Condition C (16% correct vs. 61% correct). We do not consider the production test results any further.

Table 4 shows the proportions correct on the comprehension and matching trials of the fourth and final test for both Condition C and Condition S. A 2 (condition) × 2 (test type—comprehension–matching) ANOVA of performance revealed no main effects but a

Table 3
Example of Matching Trials

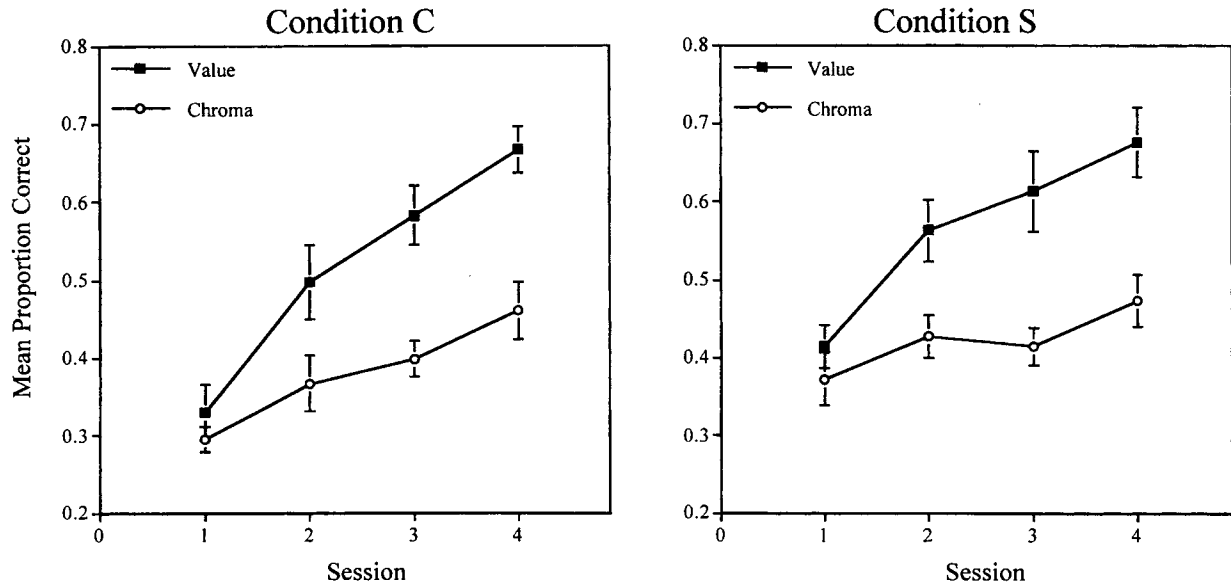|	Exemplars		Choice cards		
Condition	Exemplar 1	Exemplar 2	Correct	Incorrect	Incorrect
Color					
Value	7	7	7	3	5
Chroma	8	1	4	8	1
Size					
Value	Lightest	Lightest	Lightest	Darkest	Midbrightness
Chroma	Most saturated	Least saturated	Midsaturated	Most saturated	Least saturated

Figure 4. Mean proportion of correct value and chroma responses during the training trials for Condition Color (C) and Condition Size (S) in Study 1.

significant interaction between test type and condition, $F(1, 26) = 19.48$, $p < .01$. Thus, although participants in Condition S were trained by using the exact same procedure as the comprehension trials and participants in Condition C were trained by using a different procedure than the comprehension trials, it was the participants in Condition C who performed best on the comprehension testing trials, whereas participants in Condition S performed best on the matching test trials. Color-like training yields word comprehension before matching, whereas size-like training yields matching before word comprehension. This replicates in artificial adult learning what Sandhofer and Smith (1999) observed in the natural course of English-speaking children's learning of color and size words.

We next analyzed adult test performance in the same way that Sandhofer and Smith (1999) analyzed children's performance in their 6-month longitudinal study, by calculating the session of

Table 4

Mean Proportion Correct for Session 4 Tests in Studies 1–4

Experiment and test	Condition C: Production training	Condition S: Comprehension training
Study 1		
Comprehension	.68	.52
Matching	.41	.70
Study 2		
Comprehension	.91	.47
Matching	.75	.61
Study 3		
Comprehension	.71	.41
Matching	.43	.66
Study 4		
Comprehension	.66	.51
Matching	.45	.66

Note. C = color; S = size.

"acquisition." We defined the session of acquisition as the session in which participants responded correctly on 75% of trials with no requirement that participants maintain that level in subsequent sessions. When participants did not exceed 75% performance on any of the four sessions, we conservatively credited them with achieving 75% performance on a projected fifth session. The session of acquisition was calculated separately for two tasks: comprehension and matching. Figure 5 shows the mean session of acquisition for the comprehension and matching trials in the two training conditions. An ANOVA conducted on the session of acquisition for test type (comprehension or matching), and training condition revealed a significant interaction between test type and training condition, $F(1, 26) = 27.30$, $p < .01$, but no main effects. Thus, as indicated in Figure 5, the acquisition pattern of the comprehension and matching tests suggest that when trained in Condition C participants first learned individual words and then learned to match by the dimension; however, in Condition S participants first learned to match the dimensions and then were able to map words onto specific properties. Note that it is not the case that participants were merely delayed in learning words or abstracting dimensions in one condition but the pattern of responses is reversed for both the comprehension and matching tests in the two conditions. A production-training condition that emphasizes learning words that refer to categories is easier than (and does not promote as well) matching objects by categorical properties, but a comprehension-training condition that emphasizes learning words that refer to relations is harder than (but promotes) relationally matching objects. Thus, the learning trajectories for participants in Condition C resemble the trajectories of children learning color words, and the learning trajectories for participants in Condition S resemble children learning size words.

Although seldom occurring, some participants occasionally backtracked, that is met the criteria for acquisition at Session 3 but then did not meet criteria at Session 4, when acquiring the dimen-
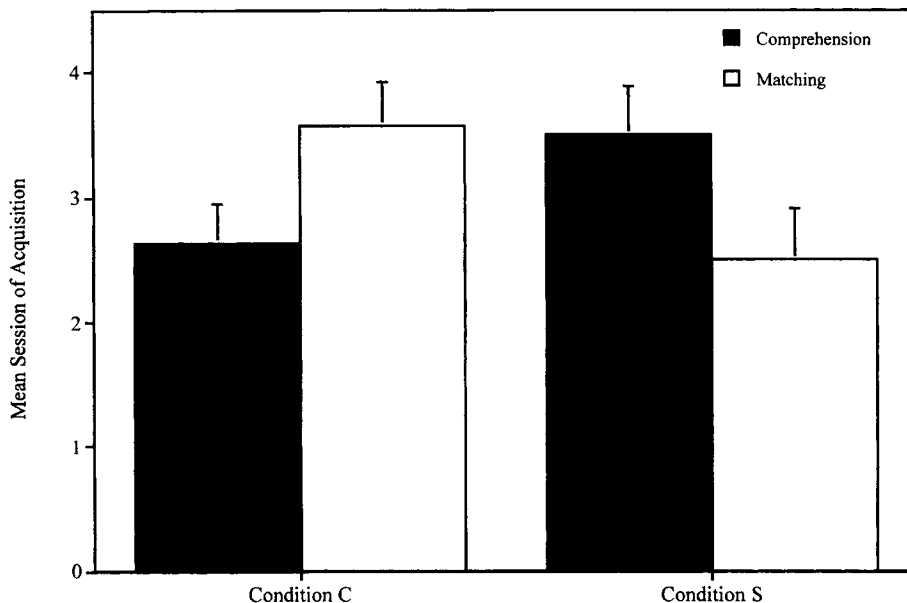
Figure 5. Mean session of acquisition in the comprehension and matching trials in Study 1. C = color; S = size.

sion words. Thus we also analyzed the participants' pattern of responses in terms of the overall correct responses. An ANOVA conducted on the test scores for test type (comprehension or matching) and training condition revealed a main effect of condition, $F(1, 14) = 13.54, p < .01$. The ANOVA also revealed a significant interaction between test type and training condition, $F(1, 110) = 9.62, p < .01$, but no main effect of test type. These results suggest participants in Condition C not only learned the dimension words before they learned to match by the dimensions, they were also more often correct in comprehending the words than in matching by the dimensions; and, participants in Condition S not only learned to abstract the dimension words before they learned the dimension words, they were also more often correct in matching the dimensions than in comprehending the dimension words.

Discussion

The question addressed in the first experiment was whether we could create the learning trajectories children commonly exhibit when learning color words by training adults to name categorically arranged values, using a category labeling procedure (Condition C), and the learning trajectories children commonly exhibit when learning size words by training participants to name relationally dependent property terms, using a stimulus selection procedure (Condition S). We succeeded in producing both sets of patterns. These results suggest that the different task demands involved in learning about dimension words, such as those for color and size, and/or the categorical–relational meaning of the dimensions are somehow responsible for creating the different developmental patterns. However, it is unclear from these results whether the training procedures, the categorical or relational nature of what is to be learned, or the interaction of the two is responsible for these learning trajectories. Studies 3 and 4 address these issues. How-

ever, before turning to these issues, in Study 2 we replicated Study 1 with a different combination of dimensions. We asked whether we can create the same learning trajectories when participants are trained with perceptually different stimuli. Thus, we asked whether the learning trajectories observed in Study 1 were due to the training procedures or the semantic and perceptual properties of the to-be-learned stimuli.

Study 2

In Study 2, we replicated the design and procedures of Study 1 but trained participants on the separable dimensions of brightness and angle degree. In this way, we addressed the issue of whether the particular perceptual dimensions to which words are mapped play a role in determining the learning trajectories. We selected the separable (by Garner's, 1974, criteria) dimensions of brightness and angle degree because the combination of these two dimensions presents participants with very different perceptual information than the integral combination of brightness and saturation used in Study 1. If the learning trajectories in Study 2 show the same patterns of acquisition as in Study 1, learning the dimension names before abstracting the dimension in the color-simulation condition and abstracting the dimension before learning the dimension names in the size-simulation condition, then the pattern of acquisition would suggest that the task structure and/or the categorical versus relational meanings of the terms strongly influence the course of learning. However, if the learning trajectories present a different pattern of acquisition than seen in Study 1, then it would suggest that the sensory and perceptual properties of specific dimensions, such as the relative ease of selective attention prior to training, matter in producing the different acquisition patterns.

Method

Participants

Twenty-eight native English-speaking undergraduates from Indiana University participated in this experiment in partial fulfillment of introductory psychology course requirements. In each condition, half of the participants were female and half were male. Each participated in four 1-hr sessions.

Materials and Design

Stimuli consisted of 1 × 1-in. squares of Coloraid gray paper pasted on a 3 × 5-in. index card with a red smiling face stamped in the lower right corner or of 1 × 2-in. pentagons of Coloraid gray paper pasted on a 3 × 5-in. index card with a blue frowning face stamped in the lower right corner. For both the Coloraid squares and the pentagons, the brightness of the gray shapes presented to participants ranged from 1 to 6 (Coloraid notation). Additionally, an angle was constructed with two thin strips of green paper and affixed to the top of each shape. The angles were either 25°, 55°, 85°, 115°, 145°, or 175°. The stimuli thus formed a 2 × 6 × 6 matrix, with each shape–face combined with every available brightness level and every angle degree. Figure 6 shows one combination of attributes.

In both conditions, the participants were required to learn words that referred to categories of or relations among brightnesses and angle degrees. In Condition C, these dimensions were referred to as value and chroma, as in Experiment 1. The shapes and the smiling or frowning faces were added as in Experiment 1, to make the selective attention task more difficult and more like that encountered by children. Variation on shape and face were irrelevant to the learning task.

In Condition C, the three categories on each dimension were defined by fixed levels. For brightness, they were low brightness (levels 5 and 6), medium brightness (levels 3 and 4), and high brightness (levels 1 and 2). For angle degree, they were small angles (25° and 55°), medium angles (85° and 115°), and large angles (145° and 175°). The particular verbal tokens assigned as the label for each category (dax, wug, melgy, zup, slee, rif) were randomly assigned. In Condition C, participants were presented with three stimulus cards on each trial, but only one was pointed to, and participants were asked to label the one card to which the experimenter pointed. Figure 7 (left) specifies the mapping of labels to stimulus categories in Condition C. Note, nine unique categories are defined by the conjunction of category labels. However, each value–chroma combination spans eight unique combinations of face–shape, angle degree, and brightness values. For example, the combination rif–wug encompasses the eight

face–angle degree–brightness combinations of Smile/25°/1, Smile/25°/2, Smile/55°/1, Smile/55°/2, Frown/25°/1, Frown/25°/2, Frown/55°/1, and Frown/55°/2.

In Condition S, three relations were defined on each dimension: lowest level present, medium level present, and highest level present. Figure 7 (right) specifies the mapping of labels to stimulus relations in Condition S. Three words referred to these relations on the brightness dimension and three referred to relations on the angle–degree dimension. For example, of the three cards simultaneously presented, the card with the lowest brightness was labeled "wug," the card with the medium brightness was labeled "dax," and the card with the highest brightness was labeled "melgy." However, the specific brightness that was low, medium, or high varied across trials. Analogously, the card with the lowest angle degree was labeled "rif," the card with the medium angle degree was labeled "slee," and the card with the highest angle degree was labeled "zup." However, the specific angles that were low, medium, or high varied across trials. During training, the three cards that were presented on every trial were arranged so that each available brightness and angle degree were presented equally often as the lowest, medium, and highest of the three cards. However, in some instances certain cards could not be presented in all three relational positions across trials. For example, values of 1 (the highest available brightness) were never presented as the low or medium brightness of the three cards. Participants were presented with three cards simultaneously and asked to select one as the correct answer.

In both conditions, participants were presented with 63 training trials at each session—seven trials from each of the nine possible brightness–angle degree label combinations (wug–rif, wug–slee, wug–zup, dax–rif, dax–slee, dax–zup, melgy–rif, melgy–slee, melgy–zup). In both conditions, nine cards, one randomly chosen from each of the brightness–angle degree label combinations of Condition C, were reserved for the testing portions of the experiment, and participants were not trained to name these nine stimuli. The order of the test trials was randomly determined.

Procedure

The procedure was identical to Study 1.

Results

Pretest

When presented with two exemplars that matched in brightness or angle degree, participants in Condition C were, prior to training,
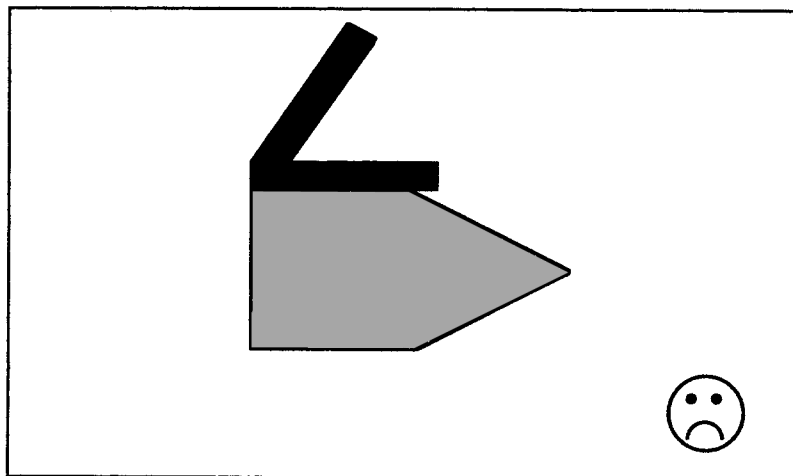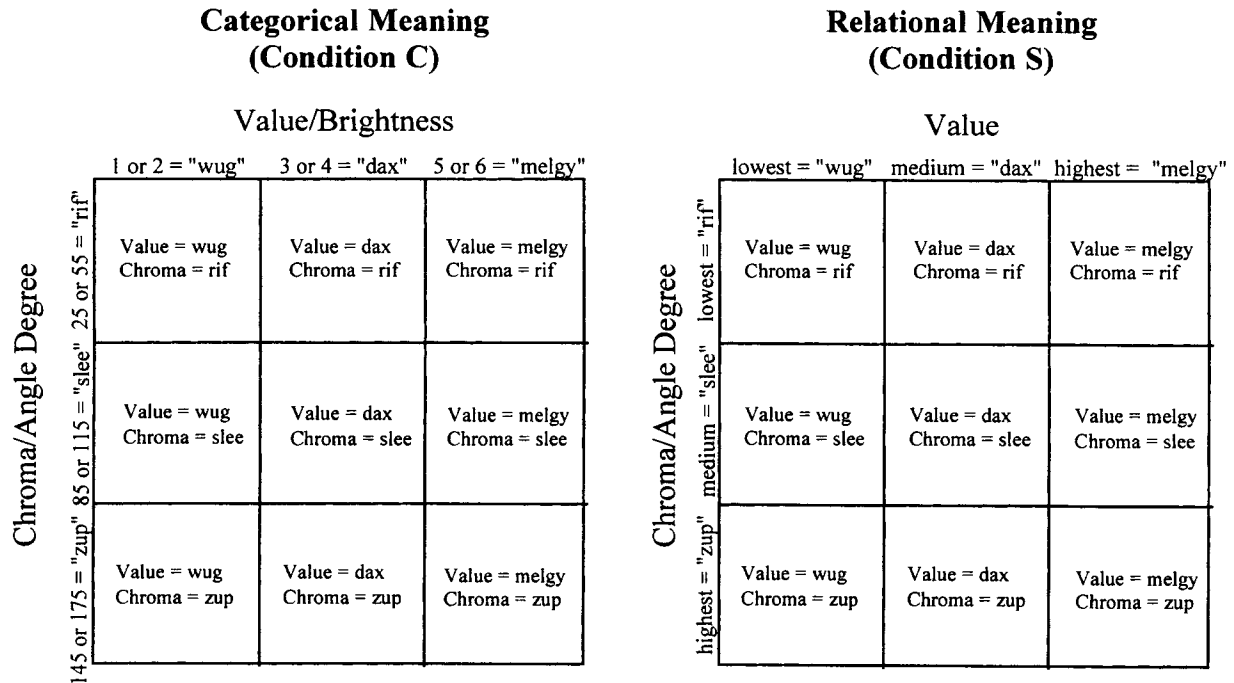

Figure 6. Example of stimuli presented in Study 2.

Categorical Meaning (Condition C)

Value/Brightness

	1 or 2 = "wug"	3 or 4 = "dax"	5 or 6 = "melgy"
25 or 55 = "rif"	Value = wug Chroma = rif	Value = dax Chroma = rif	Value = melgy Chroma = rif
85 or 115 = "slee"	Value = wug Chroma = slee	Value = dax Chroma = slee	Value = melgy Chroma = slee
145 or 175 = "zup"	Value = wug Chroma = zup	Value = dax Chroma = zup	Value = melgy Chroma = zup

Chroma/Angle Degree

Relational Meaning (Condition S)

Value

	lowest = "wug"	medium = "dax"	highest = "melgy"
lowest = "rif"	Value = wug Chroma = rif	Value = dax Chroma = rif	Value = melgy Chroma = rif
medium = "slee"	Value = wug Chroma = slee	Value = dax Chroma = slee	Value = melgy Chroma = slee
highest = "zup"	Value = wug Chroma = zup	Value = dax Chroma = zup	Value = melgy Chroma = zup

Chroma/Angle Degree

Figure 7. The mappings of labels to stimulus categories in Study 2. C = color; S = size.

able to select a third object that contained the same value, choosing the matching object on average, 54% of the time, a pattern that differed reliably from chance (chance = 33%), $t(13) = 2.38, p < .05$. However, when presented with two relationally specified examples (e.g., the lightest in their respective sets of three), participants in Condition S were unable to select the choice object that stood for the same relation within its set, choosing the relational match 26% of the time, a pattern that did not differ reliably from chance. This pattern of results tells us that without language training, the categorical matches are more obvious to participants than the relational matches.

Training

We asked how well participants learned the property terms in the two training conditions by examining the proportions of correct responses. A 2 (training condition) × 2 (dimension: brightness–angle degree) × 4 (session) ANOVA yielded a main effect of dimension, $F(1, 26) = 8.08, p < .01$, and a main effect of session, $F(3, 24) = 6.30, p < .01$, but no main effect of condition. The ANOVA also yielded significant interactions of Session × Condition, $F(3, 24) = 11.20, p < .01$, and Session × Dimension × Condition, $F(3, 24) = 14.86, p < .01$. Participants tended to learn about brightness before angle degrees, and they learned the dimension terms as a whole faster in Condition C than in Condition S. Overall, participants' patterns of responses in learning the dimension terms were similar in the two conditions, and to that of Experiment 1. The mean proportion correct for the four training trials is listed in the Appendix.

In Condition C, participants again had the opportunity to learn associations between the names for the dimensions—value (brightness) and chroma (angle degree)—and the property terms. The

evidence suggests they did so readily. Just as young children rapidly learn to answer the question "What color is it?" with a color word (even when they cannot map names to colors), adults also rapidly learned to answer the question "What chroma is it?" with one of the terms for angle degree and the question "What value is it?" with one of the terms for brightness (even though they often selected the wrong term early in training). Indeed, during the first training session, adults answered chroma questions by supplying chroma property labels and answered value questions by providing value property labels on 75% of training trials. By the end of Session 1, participants in Condition C knew which three of the six terms were chroma terms and which were value terms.

Testing

We next report performance on the tests at the end of learning. Because participants in Condition S had never been exposed to the words value or chroma in the production question "What value/chroma is this?", participants in Condition S performed considerably worse on the production test than participants in Condition C (14% correct vs. 91% correct). Again, we do not consider the results in the production task further.

Table 4 shows the proportions correct on the fourth comprehension and matching tests for both Condition C and Condition S. Participants' responses in the comprehension and matching tasks were analyzed by a 2 (condition) × 2 (test type: comprehension–matching) ANOVA. This analysis revealed a main effect of condition, $F(1, 26) = 13.65, p < .01$, no main effect of test type, and a significant interaction between test type and condition, $F(1, 26) = 4.28, p < .05$. Thus, overall participants had an easier time learning the terms and matching when the terms referred to categories rather than relations. However, a post hoc analysis con-

firmed that with these separable dimensions, just as with the integral dimensions of Experiment 1, participants in Condition C found learning the category labels easier than matching by categories, whereas participants in Condition S found matching by relations easier than mastering the relational terms, Tukey's honestly significant difference, $p < .05$.

As in Experiment 1, we defined the session of acquisition as the session in which participants responded correctly on 75% of trials with no requirement that participants maintain that level in subsequent sessions. When participants did not exceed 75% performance on any of the four trials, we conservatively credited them with achieving 75% performance on a projected fifth session. The session of acquisition was calculated separately for two tasks: comprehension and matching. Figure 7 shows the mean session of acquisition for the comprehension and matching trials in the two training conditions.

An ANOVA conducted on the session of acquisition for test type (comprehension or matching) and training condition revealed a main effect of training condition, $F(1, 26) = 15.25, p < .05$, and a significant interaction between test type and training condition, $F(1, 26) = 6.88, p < .05$, but no main effect of test type. Thus, as indicated in Figure 8, the acquisition pattern of the comprehension and matching tests suggests that when trained in Condition C participants first learned individual words and then learned to abstract the dimension and were able to learn both earlier than participants in Condition S; however, in Condition S participants first learned to abstract the dimension and then were able to map words onto specific properties but learned more slowly and with more difficulty than participants in Condition C. However, on these measures and with separable stimuli, the learning trajectories for participants in Condition C resembled the trajectories of children learning color words, and the learning trajectories for participants in Condition S resembled children learning size words.

Although occurring seldomly, some participants occasionally backtracked (e.g., met criteria at Session 3 then did not meet criteria at Session 4) when acquiring the dimension words. Thus, we also analyzed the participants' pattern of responses in terms of the overall correct responses. An ANOVA conducted on the test scores for test type (comprehension or matching) and training condition revealed a main effect of training condition, $F(1, 110) = 26.11, p < .01$, and a significant interaction between test type and training condition, $F(1, 110) = 3.99, p < .05$, but no main effect of test type. Thus, these results suggest participants in Condition C not only learned the dimension words before they learned to abstract the dimensions but they were also more often correct in comprehending the words than in abstracting the dimensions; and, participants in Condition S not only learned to abstract the dimension words before they learned the dimension words, but they were also more often correct in abstracting the dimensions than in comprehending the dimension words.

Discussion

The question addressed in the second experiment was whether in a novel dimension learning experiment with separable stimuli, could we replicate the learning trajectories seen in Study 1 and commonly exhibited by children learning about colors and sizes? We found that the learning trajectories were replicated in both Condition C and Condition S. This finding suggests that the way information is presented in a learning task and/or the categorical or relational nature of what is to be learned robustly influences how learning progresses.

However, it is important to note that the type of stimuli presented did have an effect on participants' learning as evidenced by the main effect between the two training conditions. That is, although the type of stimuli to be learned did not affect the
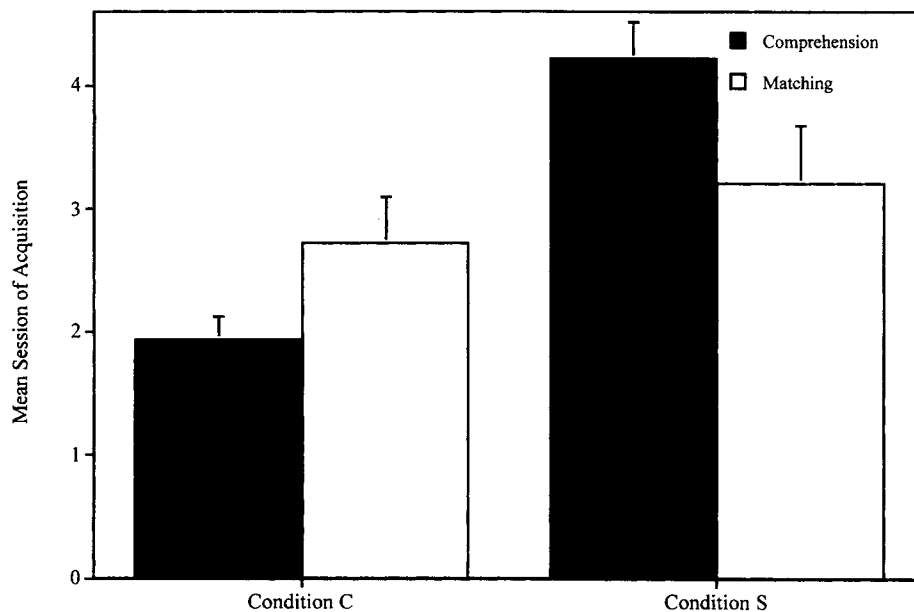

Figure 8. Mean session of acquisition in the comprehension and matching trials in Study 2. C = color; S = size.

learning trajectories for the two training conditions, it did affect the overall time it took participants to acquire the tasks in the two groups: The participants who learned separable stimuli as categories were faster than the participants who learned separable categories as relations. Moreover, in Condition C, participants in Study 2 learned to name the separable dimensions of brightness and angle degree faster than participants in Study 1, who learned the integral dimensions of brightness and saturation, $t(26) = 2.32$, $p < .05$. Surprisingly, however, participants in Condition S in Study 2 learned to name separable dimensions slower than participants in Condition S in Study 1, learning integral dimensions, $t(26) = 2.13$, $p < .05$. It is possible that the ease with which these particular dimensions could be attended to selectively hurt participants' ability to learn the relational meanings. That is, because participants could readily see the stimuli as disjunctive categories (the two that look like this are the "rif" ones) they may not have learned enough about the dimension to compare between instances.

A second question from the results of this experiment is why did participants in Condition C quickly learn to label the properties but were not immediately able to match by those properties in the nonverbal matching task? This result seems especially surprising, given that the properties were separable. One explanation is that although the dimensions were perceptually available to the participants, they still had not learned the role of words in guiding selective attention at the time of their initial success in the labeling task. Participants could succeed in the labeling task by learning a whole set of properties that are associated with a specific word. For example, instead of learning that "wug" describes a dark shape, participants may have initially learned that "wug" can be a card with either a smiling face or a frowning face on it that has a dark shape, and sometimes that shape is square.

Study 3

In Studies 1 and 2, we presented participants with training tasks designed to mimic the meaning and task structure characteristic of learning color words (i.e., categories queried in a production task) and size words (i.e., relations queried in a comprehension task). Although Experiments 1 and 2 replicated the learning trajectories children commonly present when learning colors and sizes, these studies do not disambiguate between the roles of the task structure and meaning. In Experiments 3 and 4, we examined how participants' learning trajectories are affected by both the method of training—production or comprehension—and what is to be learned—the relational or categorical meanings.

Figure 9 shows the contributions of Studies 3 and 4. In Study 3, we asked, what effect does training have in and of itself when participants are required to learn novel dimensions that describe fixed level categories? Half of the participants were asked a color-like production question (e.g., "What color is this?") and learned dimensional adjectives defined by categorically arranged values. We refer to this condition as production training. The other half of participants were asked a size-like comprehension question (e.g., "Which one is big?") and learned dimensional terms defined by categorically arranged values. We refer to this condition as comprehension training. In Study 4, we taught participants with relationally defined terms and trained them with either a color-like
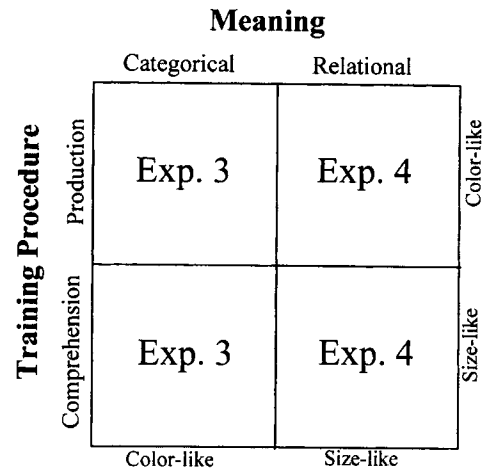

Meaning

Figure 9. The training procedures and dimension meanings tested in Studies 3 and 4. Exp. = Experiment.

question (production training) or a size-like question (comprehension training).

Method

Participants

Twenty-eight native English-speaking undergraduates from Indiana University participated in this experiment in partial fulfillment of introductory psychology course requirements. In each condition, half of the participants were female and half were male. Each participated in four 1-hr sessions.

Materials and Design

The materials were identical to those of Study 1 with two changes. First, we replaced the dimension label "value" with the novel term "velux" and replaced "chroma" with the novel term "stomin" to ensure that participants had no familiarity with the terms prior to participation in the training sessions. Second, in both conditions, participants were required to learn dimensions defined by fixed levels of brightness and saturation (equivalent to Condition C in Study 1). Thus, participants learned the very same dimension-term meanings, categorical meanings as in Condition C in Experiments 1 and 2. All that varies between the two conditions is the type of training condition: production or comprehension.

Procedure

Training. The training regimen was identical to that of Study 1. Participants were randomly assigned to one of two training conditions: production training or comprehension training.

The production-training procedures were identical to the training procedures used in Condition C of Study 1. During each trial, three cards were placed in front of the participant. The center card was pointed to by the experimenter, and the participant was asked either "What velux is this?" or "What stomin is this?" If the participant answered correctly, then the experimenter responded "You're right. That velux is wug." If the participant answered incorrectly, then the experimenter responded "No. That velux is wug."

The comprehension-training procedures were identical to the training procedures used in Condition S of Study 1. During each trial of training, three cards were placed in front of the participant, and the participant was asked to, for example, "Show me the wug one." If the participant selected

the correct card, then the experimenter responded "You're right. That's the wug one." If the participant selected the incorrect card, then the experimenter responded "No. This is the wug one," and pointed to the correct card.

Testing. The testing procedures were identical to those used in Study 1 with one important change. Recall that in Study 1 the nonverbal matching task and pretest required participants to choose a relational choice when participants were trained to name relationally dependent categories and a categorical choice when participants were trained to name categories that describe fixed values. In this study, each trial of the nonverbal matching task allowed participants to select a choice that matched relationally, a choice that matched categorically, or a foil that did not match on any dimensions. This gave us the opportunity to observe whether participants preferred one kind of match regardless of training.

Results

Pretest

The pretest results again showed that participants in the two training conditions could not match brightnesses and saturations prior to training. More specifically, when presented with two exemplars that matched by both having the same fixed value of brightness or saturation and by representing the same relation of brightness or saturation in their respective sets of three, participants in the production-training condition were, prior to training, unable to select a third object that contained the same value, choosing the categorically matching object on average 32% of the time and the relational match 25% of the time, neither of which differed from chance. Similarly, participants in the comprehension-training condition were, prior to training, unable to select a third object that contained the same value, choosing the categorically matching object on average 36% of the time, which did not differ from chance, and the relational match 26% of the time, which did not differ from chance.

Training

We next asked what participants learned during the training sessions. The mean proportion correct for the four training trials is listed in the Appendix. A 2 (training condition) × 2 (dimension: brightness–saturation) × 4 (session) ANOVA yielded a main effect of dimension, $F(1, 26) = 8.45$, $p < .01$, and a main effect of session, $F(3, 24) = 53.94$, $p < .01$. The ANOVA also yielded a significant interaction of Session × Condition, $F(3, 24) = 16.855$, $p < .01$. Overall, participants tended to learn the brightness categories before the saturation categories and learned faster in the production-training condition than in the comprehension-training condition.

However, in the production-training condition, participants had the opportunity to learn associations between the names for the dimensions—velux and stomin—and the property terms. The evidence suggests they did so readily. Just as young children rapidly learn to answer the question "What color is it?" with a color word (even when they cannot map names to colors) adults also rapidly learned to answer the question "What stomin is it?" with one of the "stomin" terms and the question "What velux is it?" with one of the "velux" terms (even though they often selected the wrong term early in training). Indeed, during the first training session, adults answered stomin questions by supplying stomin property labels and answered velux questions by providing velux property labels on 69% of training trials. By the end of Session 1, participants in

the production-training condition knew which three of the six terms were stomin terms and which were velux terms.

Testing

We first examined participants' responses on the matching trials and asked whether participants were more likely to choose categorical or relational matches at each session. In both conditions, participants' selections of relational matches never exceeded that expected by chance at any session. We therefore counted only categorical matches as correct responses in the matching trials.

We first report performance on the tests at the end of learning. Because participants in the comprehension-training condition had never been exposed to the words *value* and *chroma* in the production question, "What value/chroma is this?" participants in this condition performed considerably worse on the production test than participants in the production-training condition (11% correct vs. 70% correct). We did not consider these production test results any further. Table 4 shows the proportions correct on the comprehension and matching trials of the fourth test for both the production-training and comprehension-training conditions. We conducted a 2 (condition) × 2 (test type: comprehension–matching) ANOVA. This analysis revealed no main effects, but a significant interaction between test type and condition, $F(1, 26) = 19.21, p < .01$. Thus, participants in the production-training condition found learning the category labels easier than matching by categories, whereas participants in the comprehension-training condition found matching by categories easier than mastering the category labels. This difference holds despite the fact that participants in the two conditions had to learn the same meanings.

As in Experiments 1 and 2, we defined the session of acquisition as the session in which participants responded correctly on 75% of trials with no requirement that participants maintain that level in subsequent sessions. When participants did not exceed 75% performance on any of the four trials, we conservatively credited them with achieving 75% performance on a projected fifth session. The session of acquisition was calculated separately for two tasks: comprehension and matching. Figure 10 shows the mean session of acquisition for the comprehension and matching trials in the two training conditions. An ANOVA conducted on the session of acquisition for test type (comprehension or matching) and training condition revealed a significant interaction between test type and training condition, $F(1, 26) = 5.60, p < .05$, but no main effects. Thus, as indicated in Figure 10, the acquisition pattern of the comprehension and matching tests suggest that in the production-training condition participants again first learned individual words and then learned to match by the dimension; however, in the comprehension-training condition participants first learned to match the dimension and then were able to map words onto specific properties. This is so despite the fact that the terms in both conditions mapped to the same meaning. The clear implication from these results is that the training—production versus comprehension—not the meaning of the to-be-learned terms determines the developmental pattern in the comprehension and matching tasks.

Again we also analyzed the participants' pattern of responses in terms of the overall correct responses. An ANOVA conducted on the test scores for test type (comprehension or matching) and training condition revealed a significant interaction between test
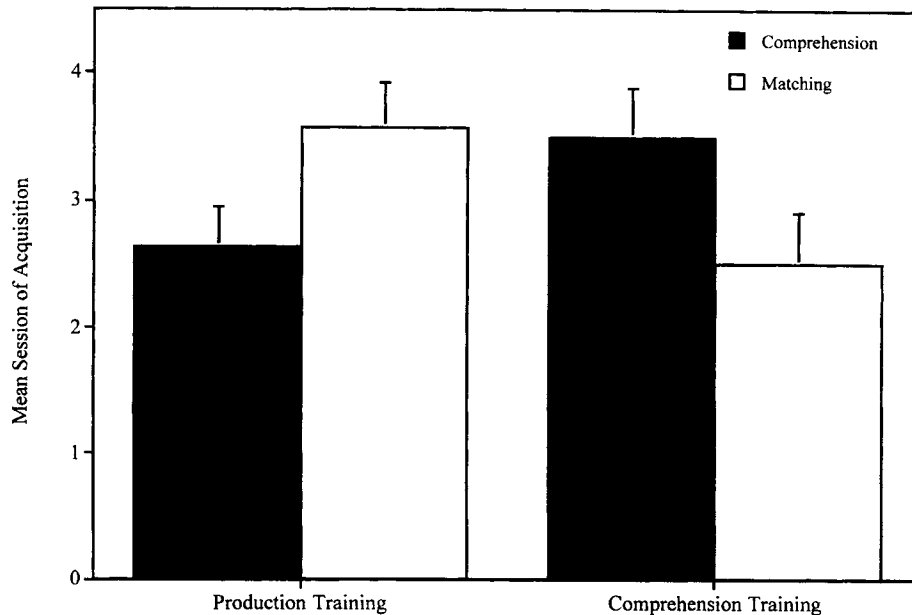
Figure 10. Mean session of acquisition in the comprehension and matching trials in Study 3.

type and training condition, $F(1, 110) = 6.96, p < .01$, but no main effects. Thus, participants in the production-training condition not only learned the dimension words before they learned to match by the dimensions, but they were also more often correct in comprehending the words than in matching the dimensions; and participants in the comprehension-training condition not only learned to match the dimensions before they learned the dimension words, but they were also more often correct in matching the dimensions than in comprehending the dimension words.

Discussion

The results strongly suggest that it is the training procedure not the meanings that determine the learning pattern. In the next experiment, we sought further confirmation of this idea. Again, we contrasted production training in which the learner's attention is focused on one object and the learner must produce a name for the queried property versus comprehension training in which participants are presented with a property term and must select between alternatives. However, in Experiment 4, the meanings in both conditions are relational. If the training procedure—production versus comprehension—is all that matters, then we should replicate the patterns of Experiments 1, 2, and 3. Word comprehension should precede matching, given production training, but success in matching should precede word comprehension, given comprehension training.

Study 4

Method

Participants

Twenty-eight native English-speaking undergraduates from Indiana University participated in this experiment in partial fulfillment of introductory

psychology course requirements. In each condition, half of the participants were female and half were male. Each participated in four 1-hr sessions.

Materials and Design

The materials were identical to those of Study 3 with one change. In both conditions, participants were required to learn dimensions defined by relationally dependent meanings of brightness and saturation (equivalent to Condition S in Study 1). Thus, participants learned the very same dimension-term meanings, relational meaning as in Condition S. All that varies between the two conditions is the type of training condition, which we refer to as the production-training and comprehension-training conditions.

Procedure

The procedure was identical to that in Study 3.

Results

Pretest

The pretest results again showed that participants in the two training conditions could not match brightnesses and saturations prior to training. Participants in the production-training condition were, prior to training, unable to select a third object that contained the same value, choosing the categorically matching object on average 32% of the time and the relational match 30% of the time, neither of which differed from chance. Similarly, participants in the comprehension-training condition were, prior to training, unable to select a third object that contained the same value, choosing the categorically matching object on average 39% of the time, which did not differ from chance, and the relational match 34% of the time, neither of which differed from chance.

Training

We next asked what participants learned during the training sessions. The mean proportion correct for the four training trials is listed in the Appendix. A 2 (training condition) × 2 (dimension: brightness–saturation) × 4 (session) ANOVA yielded a main effect of dimension, $F(1, 26) = 10.50, p < .01$, and a main effect of session, $F(3, 24) = 13.02, p < .01$, but no main effect of condition. The ANOVA also yielded an interaction of Dimension × Session, $F(3, 24) = 6.55, p < .01$. Again participants tended to learn brightness terms before saturation terms.

Again in the production-training condition, participants had the opportunity to learn associations between the names for the dimensions—velux and stomin—and the property terms. Again, the evidence indicates they did so. During the first training session, adults answered stomin questions by supplying stomin property labels and answered velux questions by providing velux property labels on 72% of training trials. By the end of Session 1, participants in the production-training condition knew which three of the six terms were stomin terms and which were velux terms.

Testing

We first examined participants' responses on the matching trials and asked whether participants were more inclined to choose categorical or relational matches at each session. In both conditions, participants' selections of categorical matches never exceeded that expected by chance at any session. We therefore only scored relational matches as correct responses in the matching trials.

Participants in the comprehension-training condition had never been exposed to the words velux and stomin during training, and thus they performed poorly on all production tests (8% correct),

whereas participants in the comprehension-training condition performed well (59% correct).

Table 4 shows the critical results: the proportions correct on the comprehension and matching trials on the fourth and final test. We conducted a 2 (condition) × 2 (test type: comprehension–matching) ANOVA. This analysis revealed no main effects, but it did reveal a significant interaction between test type and condition, $F(1, 26) = 18.18, p < .01$. Again, participants in the production-training condition found that learning the category labels was easier than matching by categories, whereas participants in the comprehension-training condition found that matching was easier than mastering the labels. This pattern is identical to that of Experiment 3: Production training fosters learning the words before being able to match objects on the property, and comprehension training fosters learning to match before learning the words, even given relational meanings.

We defined the session of acquisition as in the earlier experiments. Figure 11 shows the patterns of acquisition. An ANOVA conducted on the session of acquisition for test type (comprehension or matching) and training condition revealed a significant interaction between test type and training condition, $F(1, 26) = 29.80, p < .01$, but revealed no main effects. Again, production training led to success in the word-comprehension task prior to success in the matching task, but comprehension training led to the opposite pattern, and this was so despite that the meanings to be learned and the match deemed correct were relational in both cases.

This same pattern emerged in the analysis of overall correct responses. An ANOVA conducted on the test scores for test type (comprehension or matching) and training condition revealed a significant interaction between test type and training condition, $F(1, 110) = 27.62, p < .01$, but revealed no main effects. Thus, participants in the production-training condition not only learned
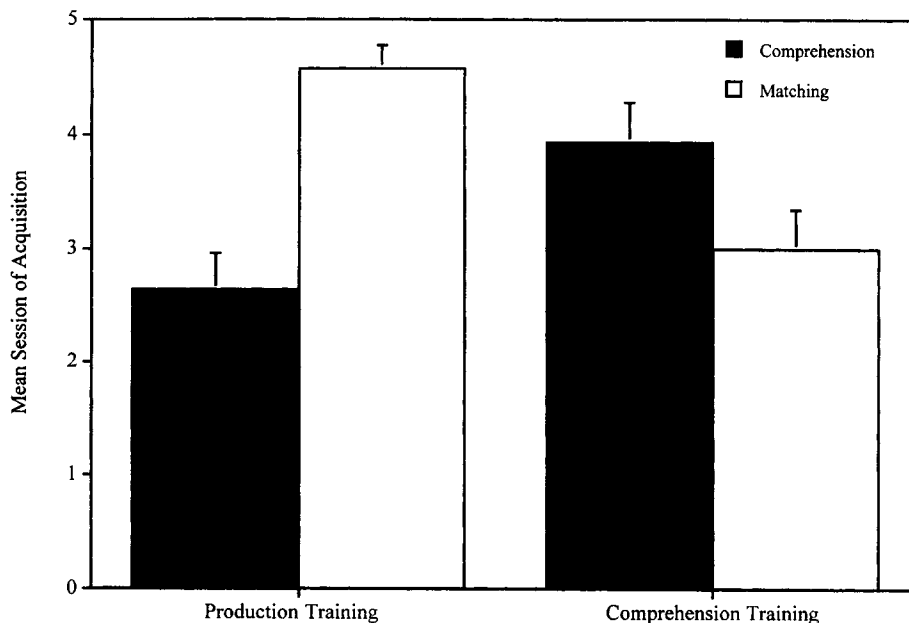

Figure 11.   Mean session of acquisition in the comprehension and matching trials in Study 4.

the dimension words before they learned to match by the dimensions, but they were also more often correct in comprehending the words than in matching the dimensions. In addition, participants in the comprehension-training condition not only learned to match the dimensions before they learned the dimension words, but they were also more often correct in matching the dimensions than in comprehending the dimension words.

General Discussion

The results of these studies strongly suggest that the structure of the training task is a potent force on whether word learning precedes attention to dimensional similarities and differences in a nonlinguistic matching task or whether attention to dimensional similarities and differences precedes word knowledge. The developmental relation between children's performances in these two kinds of tasks has been a focus of research interest because the ordering seems relevant to the issue of whether dimensional concepts precede word learning or whether they are constructed through word learning. The finding that children can succeed in nonlinguistic matching tasks but fail in word-comprehension tasks has been interpreted as indicating available concepts but lack of word knowledge. In contrast, the finding that children succeed in nonlinguistic matching tasks only at the same time or after acquisition of the words has been interpreted as consistent with the idea that learning dimension words creates dimensional concepts. The present results raise complex questions about these standard interpretations, in that they suggest that the training regimen, not the underlying meanings to be learned, determines the relative primacy of success in the nonlinguistic and linguistic tasks.

We pursued the implications of the present results in three steps. We considered first what the word comprehension and matching tasks are measuring. We considered second what it is that the training tasks are teaching that lead to earlier versus later attention to dimensional similarities and nonsimilarities in the nonlinguistic matching tasks. Finally, we returned to the developmental issues with which we began: why children's learning of color and size words exhibit such different developmental patterns and what this learning means for the nature of dimensional word learning and developmental process more generally.

Matching and Word Comprehension

Our reasoning in comparing adults' performances in the word-comprehension and nonlinguistic matching tasks is the same as in our and others' developmental studies of these same issues. Are learners attending to the relevant properties and relations for learning the words? The implicit assumption behind this question and the comparison of performance in nonlinguistic and linguistic tasks is that they measure somewhat separate kinds of knowledge—knowledge about perceptual properties and relations versus knowledge about the words that refer to those properties and relations. At a surface level, this reasoning still seems right. In the nonlinguistic matching task, participants are presented with a pair of objects that differ in many ways and are alike in only one way. The participant's task is to find another object that matches in the same way. Success in this task would seem to require a perceptual analysis of similarities and differences into dimensional kinds and attention to specific similarities and differences. Participants do

not, in principle, need to know the labels for the specific properties they are matching nor for the dimensions.

However, the present results make clear that experiences in learning dimension words drive performance changes in this non-linguistic matching task as well as in the word-comprehension task. Over the course of training, participants trained with both the word-comprehension and word-production formats become better able to make the matches in the nonlinguistic task. At one level, this is not the least bit surprising. Goldstone (1994), among others (Kruschke, 1992; Nosofsky, 1986; Logan, Lively, & Pisoni, 1991; Lane, 1965), has shown that learning category labels increases attention to and discrimination among the dimensions relevant for categorization. In the same way, then, having to attend to the properties and dimensions in the word-learning tasks of the present experiments increases attention to those same properties and dimensions in the nonlinguistic task.

What is surprising is that success in the matching task precedes word knowledge by one training regimen and follows it by another. In general, a finding that participants succeed in one task before another could mean only that one task is a more sensitive measure of the same learning or more generally easier. But this is not (at least not simply) the case here. The very same tasks differ in their order of acquisition for different sets of learners.

One reason that the order of success in the matching and word-comprehension tasks could be different under the two training regimens is that the matching task is solved very differently and is not the same task for the two groups of learners. One way that one might solve the matching task is illustrated in Figure 12. The participant looks at the two exemplar objects and perceives the matching property and searches for another object that matches in the same way. The task solved in this way is a perceptual comparison task. A second way that one might solve the task is also illustrated in Figure 12. The participant might look at one of the objects in the exemplar pair and note that it can be labeled by the property term "melgy" and then look at the second object in the exemplar pair and note that it can also be labeled by the property term "melgy." This participant might then search for another object from the choice set that can also be called "melgy." Solved in this way, the matching task is a lexical categorization task.

The possibility that participants, given production training,
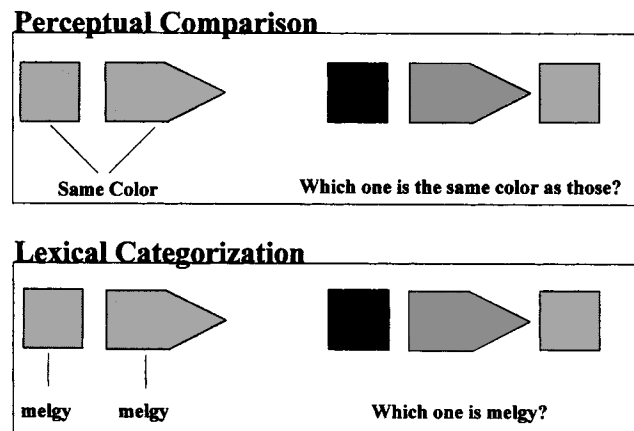

Figure 12. Two potential ways to solve the nonlinguistic matching task.

solved the supposedly nonlinguistic task through lexical categorization is consistent with the fact that they learned the words before they successfully matched objects and also with the overt labeling behavior of some participants in the matching task. That is, some participants who had received production training overtly labeled the objects with the experimental terms in the matching task, and they explicitly sought a choice object that could also be so labeled. Presumably, what some participants in the production-training conditions did overtly, others did covertly. Intriguingly, no participants in the comprehension-training conditions were observed to overtly solve the matching task in this way. Thus, for some learners, the word-comprehension and matching tasks may have both measured word knowledge. And for different learners—those receiving production training versus comprehension training—the nonlinguistic task may have measured different underlying knowledge.

All in all, these results suggest caution in making inferences about nonlinguistic versus linguistic competencies. The results suggest that the words one is learning—and the way they are taught—influence performances in nonlinguistic tasks. This conclusion has precedence in the developmental literature. In their cross-linguistic studies of English and Korean spatial and motion terms, Choi and Bowerman (1991) and Bowerman (1996) found evidence of language-specific influence from the earliest stages of learning, long before the child had complete command of the linguistic terms.

Comprehension and Production Training

What are the two training regimens teaching that leads to the different patterns of learning—patterns that seem completely determined by the training and not by the specific dimensions nor by the categorical versus relational meanings of the terms? The key difference may be one of a training regimen that focuses attention on a single object and the associated word versus one which focuses attention on similarities and differences between objects in the process of linking them to words.

In the production-training task, participants' goal on each trial is to find the word that labels one designated object. Selective attention to a specific property of that object may help in this mapping, and thus it is to be expected that participants will learn to selectively attend to the labeled properties. However, given that selective attention is trained through the linking of an object to a word, selective attention itself may be limited to and only operate through lexical categorization. Thus, learners may only know that two things match by knowing that they are "melgy."

In contrast, under the comprehension-training procedure, the participants' attention is necessarily focused on several objects and the similarities and differences among them because the learner has to find the one object of those to which the word refers. Selective attention to specific properties and to dimensions will thus be encouraged both by the comparison of the choice objects and by the linking of the chosen object to the label. Because selective attention to similarities and differences among objects is on each trial, prior to the mapping of the selected object to the word, it may emerge both prior to and independently of processes of lexical categorization.

Put simply, we are suggesting that participants narrowly learn what they are taught. In the production-training procedure, the links between an object and a label are emphasized, and participants primarily learn to link objects to labels and only secondarily to selectively attend to dimensional similarities. In the comprehension-training procedure, the similarities and differences between objects are emphasized, and thus participants primarily learn to selectively attend to the similarities between objects and secondarily to map objects to labels.

The idea that training regimens that promote categorization versus comparison yield different kinds of learning has precedence in both the developmental and adult literatures (Gentner & Markman, 1994; Gentner & Namy, 2000; Medin, Goldstone, & Gentner, 1993; Spalding & Ross, 1995; Waxman & Klibanoff, 2000.) The developmental studies, in particular, have suggested that opportunities to compare two perceptually present objects enhance children's learning about dimensions. For example, Kotovsky and Gentner (1996) reported that preschool children who compared pictures during training were better able to notice cross-dimensional correspondences based on the same relations than were control children who did not receive comparison training. Analogously, in a microgenetic training study, Namy, Smith, and Gershkoff-Stowe (1997) compared the classificatory development of 18-month olds who were trained in a task in which they compared instances of two simultaneously presented categories versus that of children who were trained by interacting with instances of each category separately. The children who had the opportunity to compare instances of each category advanced to a higher level of exhaustive classification as a result of their experiences than did the children trained separately on the two categories. Similarly, Yamauchi and Markman (2000) showed that making comparisons during learning can ease adults' ability to find a correct abstract description of stimulus dimensions when the abstract feature values differ in their specific instantiations. And, Bowdle and Gentner (1997) showed that adults' analogical problem solving was enhanced by explicit comparison of examples rather than by separate experiences with them.

Comprehension tasks contain within their structure comparison training, and thus like comparison more generally, they may enhance learners' discovery of relations. This is a new insight into the difference between comprehension and production tasks. Comprehension and production tasks are usually discussed solely under the rubric of recognition versus recall with the idea being that they involve the same processes but that comprehension is easier than production. The present results and the other research on comparison versus categorization suggest a deeper distinction: The two tasks engage and thus teach different processes with different outcomes.

Color and Size Learning and Developmental Process

There is a large literature on adult category learning that demonstrates—in different ways than those shown here—how the structure of the learning task affects what is ultimately learned. For example, (a) the number, range, and variability of instances influence whether and what kinds of category prototypes are formed (J. D. Smith & Minda, 1998; Posner & Keele, 1968); (b) the way features are grouped together and whether they are presented in a correlated manner influences what participants learn about them (Tversky, 1977; Wattenmaker et al., 1986); and (c) the order with which the same categorization tasks are taught influences what is

learned from each task (Schyns et al., 1998; Schyns & Rodet, 1997). In addition, a growing literature documents how category learning depends on the goals of the learner and the nature of the training tasks. The learner's goals affect how stimulus objects are perceived and thus what is learned (e.g., Lamberts, 1994; Ward & Becker, 1992), the learner's view of the task as prediction versus categorization affects the nature of the categories formed (Waldmann & Holyoak, 1992; Yamauchi & Markman, 1998), and participants use of a category alters feature weightings from previous learning (Ross, 1997). All these experimental studies are like the present one in showing that the structure of the learning task is itself a major force on what is learned.

This literature, however, is generally considered as separate from questions about developmental process and developmental mechanism, that is, as not germane to questions of the processes that turn the cognitive abilities of 1-year-olds into 2-year-olds and those of 2-year-olds into those of 5-year-olds. The present results suggest, however, that these findings about the role of task structure in adult category learning may well be relevant to understanding cognitive development. That is, the character of cognitive development may depend intimately on the learning tasks in which children find themselves, and it may depend just as intimately on those tasks and their structure as they do on the maturational levels of children or the meanings and concepts to be learned.

The present results strongly suggest that this is the case for size and color-word learning. Previous research with children showed that when they learn color words, children mastered the words before they successfully compared objects by color in nonlinguistic tasks. However, previous research also showed that when learning size terms, children successfully attended to the similarities and differences in size before they mastered the words. We simulated these patterns in adults by mimicking the training regimens that seemed characteristic of color learning on the one hand and characteristic of size learning on the other. Also, we found that which pattern emerged—mastery of the words before success in the matching or success in the match task before mastery of the words—depended only on the training regimen and not on the meanings to be learned. The clear implication is that what makes the developmental pattern in children's learning of size terms and what makes the developmental pattern in children's learning of color terms is the way they are taught. Although the present results only imply this and do not conclusively show it, the empirical prediction that follows is clear: If we were to teach children color words by using comprehension training that promoted comparison, then we should find a pattern like that typically characteristic for learning size terms, and if we teach children size terms by using production training that promoted categorization, then we should find a pattern typically characteristic for learning color terms.

In conclusion, the results present a new way to think about the role of language in the development of thought. The way language is used and taught differentially engages processes of categorization and comparison and, thus, the stages through which learning progresses and perhaps its ultimate endpoint. Thus, understanding how the learning task determines what is learned should lead in the long term to a deeper understanding of similarities and differences in the developmental patterns of different concepts, different individuals, and perhaps different groups of people.

References

Adams, R. J. (1989). Newborns' discrimination amid mid- and long-wave stimuli. *Journal of Experimental Child Psychology, 47*, 130–141.

Andrick, G. R., & Tager-Flusberg, H. (1986). The acquisition of colour terms. *Journal of Child Language, 13*, 119–134.

Backscheider, A. G., & Shatz, M. (1993, April). *Children's acquisition of the lexical domain of color*. Twenty-ninth annual Proceedings of the Chicago Linguistic Society, Chicago.

Bartlett, E. J. (1978). The acquisition of the meaning of colour terms: A study of lexical development. In R. N. Campbell & P. T. Smith (Eds.), *Recent advances in the psychology of language: Language development and mother-child interaction* (pp. 89–108). New York: Plenum.

Bornstein, M. H. (1985). On the development of color naming in young children: Data and theory. *Brain and Language, 26*, 72–93.

Bowdle, B. F., & Gentner, D. (1997). Informativity and asymmetry in comparisons. *Cognitive Psychology, 34*, 244–286.

Bowerman, M. (1996). Learning how to structure space for language: A crosslinguistic perspective. In P. Bloom & M. A. Peterson (Eds.), *Language and space. Language, speech, and communication* (pp. 385–436). Cambridge, MA: MIT Press.

Brown, A. M., & Teller, D. Y. (1989). Chromatic opponency in 3-month-old human infants. *Vision Research, 29*, 37–45.

Carey, S. (1982). Semantic development: State of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 347–389). Cambridge, England: Cambridge University Press.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition, 41*, 83–121.

Clark, H. H. (1970). The primitive nature of children's relational concepts. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 269–278). New York: Wiley. .

Cruse, D. A. (1977). A note on the learning of color names. *Journal of Child Language, 4*, 305–311.

Darwin, C. H. (1877). A biographical sketch of a young infant. *Kosmos, 1*, 367–376.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Gasser, M., & Smith, L. B. (1998). Learning nouns and adjectives: A connectionist account. *Language and Cognitive Processes, 13*, 269–306.

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Sciences, 5*, 152–158.

Gentner, D., & Namy, L. L. (2000). Comparison in the development of categories. *Cognitive Development, 14*, 487–513.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 135–176.

Gitterman, D., & Johnston, J. R. (1983). Talking about comparisons: A study of young children's comparative adjective usage. *Journal of Child Language, 10*, 605–621.

Gleitman, L. R., & Wanner, E. (1988). Current issues in language learning. In M. H. Bornstein & M. Lamb (Eds.), *Developmental psychology: An advanced textbook* (pp. 297–356). Hillsdale, NJ: Erlbaum.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*, 178–200.

Heider, E. R. (1971). "Focal" color areas and the development of color names. *Developmental Psychology, 4*, 447–455.

Istomina, Z. M. (1963). Perception and naming of color in early childhood. *Soviet Psychology and Psychiatry, 1*, 37–45.

Kendler, H. H., & Kendler, T. S. (1961, November 17). Effect of verbalization on reversal shifts in children. *Science, 134*, 1619–1620.

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

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

Lamberts, K. (1994). Flexible tuning of similarity in exemplar-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1003–1021.

Landau, B., & Gleitman, L. R. (1985). *Language and experience.* Cambridge, MA: Harvard University Press.

Lane, H. (1965). The motor theory of speech perception: A critical review. *Psychological Review, 72*, 275–309.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874–886.

Markman, E. M. (1989). *Categorization and naming in children.* Cambridge, MA: MIT Press.

Martin, J. E. (1969). Semantic determinants of preferred adjective order. *Journal of Verbal Learning and Verbal Behavior, 8*, 697–704.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254–278.

Namy, L. L., Smith, L. B., & Gershkoff-Stowe, L. (1997). Young children's discovery of spatial classifications. *Cognitive Development, 12*, 163–184.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*(1–2, Serial No. 149).

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

Park, S., Tsukagoshi, K., & Landau, B. (April, 1985). *Young children's mis-naming of colors.* Paper presented at the Society for Research in Child Development, Toronto, Ontario, Canada.

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.

Ross, B. H. (1997). The use of categories affects classification. *Journal of Memory and Language, 37*, 240–267.

Sandhofer, C. (2001). Structure in parents' input: Effects of categorization versus comparison. In A. H. Do, L. Dominguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development: Vol. 2* (pp. 657–667). Sommersville, MA: Cascadilla Press.

Sandhofer, C., & Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology, 35*, 668–679.

Schyns, P., Goldstone, R. L., & Thibault, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21*, 1–54.

Schyns, P., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 681–696.

Sera, M., & Smith, L. B. (1987). Big and little: "Nominal" and relative uses. *Cognitive Development, 2*, 89–111.

Shepard, R., & Cooper, L. (1992). Representation of colors in the blind, color-blind, and normally sighted. *Psychological Sciences, 3*, 97–104.

Slater, A., & Johnson, S. P. (1998). Visual sensory and perceptual abilities of the newborn: Beyond the blooming, buzzing confusion. In F. Simion & G. Butterworth (Eds.),. *The development of sensory, motor and cognitive capacities in early infancy: From perception to cognition* (pp. 121–141). Hove, England: Erlbaum.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1411–1436.

Smith, L. B. (1989). From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogy* (pp. 146–178). New York: Cambridge University Press.

Smith, L. B., Gasser, M., & Sandhofer, C. M. (1997). Learning to talk about the properties of objects: A network model of the development of dimensions. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.), *Perceptual learning, 36, The psychology of learning and motivation* (pp. 219–255). San Diego, CA: Academic Press.

Smith, L. B., & Sera, M. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology, 24*, 99–142.

Soja, N. N. (1994). Young children's concept of color and its relation to the acquisition of color words. *Child Development, 65*, 918–937.

Spalding, T. L., & Ross, B. H. (1995). Comparison-based learning: Effects of comparing instances during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1251–1263.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Vendler, Z. (1968). *Adjectives and nominalizations.* The Hague, the Netherlands: Morton Publishers.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*, 222–236.

Ward, T. B., & Becker, A. H. (1992). Learning categories with and without trying: Does it make a difference? In B. Burns (Ed.), *Percepts, concepts and categories* (pp. 451–491). Amsterdam: Elsevier Science.

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.

Waxman, S. R., & Gelman, R. (1986). Preschoolers' use of superordinate level relations in classification and language. *Cognitive Development, 1*, 139–156.

Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology, 36*, 571–581.

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language, 39*, 124–148.

Yamauchi, T., & Markman, A. B. (2000). Learning categories composed of varying instances: The effect of classification, inference, and structural alignment. *Memory and Cognition, 28*, 64–78.

Appendix

Mean Proportion Correct in the Four Training Sessions and Two Training
Conditions for Studies 1–4

Experiment and session	Condition C: Production training		Condition S: Comprehension training	
	Value	Chroma	Value	Chroma
Study 1				
1	.33	.30	.42	.35
2	.50	.37	.56	.41
3	.58	.40	.61	.42
4	.69	.50	.68	.48
Study 2				
1	.28	.27	.38	.43
2	.48	.54	.46	.49
3	.66	.74	.51	.54
4	.75	.80	.55	.64
Study 3				
1	.67	.50	.48	.46
2	.65	.51	.46	.44
3	.27	.26	.42	.33
4	.50	.26	.41	.37
Study 4				
1	.30	.28	.42	.39
2	.48	.38	.53	.37
3	.58	.46	.58	.41
4	.64	.51	.61	.42

Note. C = color; S = size.