The Journal of General Psychology, 2000, *127*(3), 279–326

Adults' Acquisition of Novel Dimension Words: Creating a Semantic Congruity Effect

BRIGETTE OLIVER RYALLS
Department of Psychology
University of Nebraska at Omaha

LINDA B. SMITH
Department of Psychology
Indiana University

ABSTRACT. The semantic congruity effect is exhibited when adults are asked to compare pairs of items from a series, and their response is faster when the direction of the comparison coincides with the location of the stimuli in the series. For example, people are faster at picking the bigger of 2 big items than the littler of 2 big items. In the 4 experiments presented, adults were taught new dimensional adjectives (*mal/ler* and *borg/er*). Characteristics of the learning situation, such as the nature of the stimulus series and the relative frequency of labeling, were varied. Results revealed that the participants who learned the relative meaning of the artificial dimensional adjectives also formed categories and developed a semantic congruity effect regardless of the characteristics of training. These findings have important implications for our understanding of adult acquisition of novel relational words, the relationship between learning such words and categorization, and the explanations of the semantic congruity effect.

IN A NOW CLASSIC PAPER, Bierswisch (1967) suggested that the study of dimensional adjectives merited special attention for what they could tell us about the properties of mind. He argued that the structure of dimensional adjectives "did not represent the properties of the surrounding world in the broadest sense, but rather certain deep seated, innate properties which determined the way in which the universe is conceived, adapted, and worked on" (p. 3). In the present article we, like Bierswisch, seek insights into cognition by studying dimensional adjectives. Unlike Bierswisch, however, we seek to understand not the "universal" aspects of the linguistic structure of dimensional adjectives but an apparently universal aspect of the processing of dimensional adjectives. In addition, unlike Bierswisch, we seek an understanding not of the innate properties of the human mind but of the consequences of learning dimensional adjectives.

Our ideas derive from the joint consideration of two literatures that are not generally considered together, despite similar experimental phenomena. These literatures concern (a) adult comparative judgments and (b) the acquisition of comparative dimensional adjectives by young children. The similar phenomena are the *semantic congruity effect* found in adult comparative judgments and the *endpoint* and *semantic congruity effects* found in children's early comprehension of dimensional adjectives. As a background to the empirical work reported here, first, we review findings on the semantic congruity effect in adult comparative judgments; second, we review findings on similar effects in children's acquisition of dimensional adjectives; and third, we present the implications suggested by the joint consideration of these two literatures.

Adult Comparative Judgments

In comparative judgment tasks adults produce a pattern of responses known as the semantic congruity effect, which is the finding that adults are faster to compare objects drawn from a series when the direction of comparison is congruent with the location of the stimuli on the continuum. For example, when asked to make judgments about the sizes of animals, adults are faster at choosing the larger of two relatively large animals (e.g., elephant versus hippo) than at choosing the smaller of two relatively large animals. Conversely, adults are faster at choosing the smaller of two small animals (e.g., hamster versus gerbil) than the larger of two relatively small animals (Banks & Flora, 1977).

The semantic congruity effect is a general and robust phenomenon in adult judgments. Since first reported by Shipley, Coffin, and Hadsell (1945) in their investigation of color preference, congruity effects in reaction time have been found to characterize perceptual judgments such as those of pitch, brightness (Audley & Wallis, 1964; Wallis & Audley, 1964), loudness (Holyoak & Patterson, 1981), length, weight, and horizontal extent (Petrusic, 1992; Petrusic & Baranski, 1989b). The effect has been found in comparative judgments involving symbolic stimuli or conceptual information including height (Banks, Clark, & Lucy, 1975), intelligence (Banks & Flora, 1977), number magnitude (Holyoak, 1978), and size (Banks & Flora, 1977; Banks, White, Sturgill, & Mermelstein, 1983; Cech, 1995; Cech, Shoben, & Love, 1990). And finally, it has been found in judgments involving more complex relations such as relative age (Ellis, 1972), probability of events (Marks, 1972), and racial identity (Friend, 1973). In addi-

---

Portions of this research were presented at the annual meeting of the Psychonomics Society in New Orleans, LA, 1990. The completion of this research was supported in part by PHS Grant Number 1 F31 MH10403-01 SRCM-C.

Address correspondence to Brigette Oliver Ryalls, Department of Psychology, University of Nebraska at Omaha, Omaha, NE 68182-0274; brigette_ryalls@unomaha.edu (e-mail).

tion to these robust reaction time effects, Petrusic and Baranski (1989a, 1989b; Petrusic, 1992) have found the semantic congruity effect in adults' confidence judgments. However, although predicted by many theories of the comparison process, the semantic congruity effect in the accuracy of adults' comparative judgments has been elusive and documented only in a single report (Petrusic, 1992).

The semantic congruity effect in adults does not reflect a stable partitioning of dimensions into set categories in long-term memory because the effect is context specific. If presented with a series ranging from rabbit to elephant, adults are faster to choose rabbit as smaller than fox than to choose fox as larger than rabbit. In contrast, when presented with a continuum ranging in size from flea to fox, adults are faster to choose fox as larger than rabbit than they are to choose rabbit as smaller than fox. The contextual determination of the semantic congruity effect is so fast in adult comparative judgments that if the range of the stimuli is altered within a single testing session the congruity effect "moves" to fit the altered continuum (Cech & Shoben, 1985; see also Petrusic & Baranski, 1989a).

Numerous models have been proposed to account for the semantic congruity effect assuming either analogical representations, such as Holyoak's reference point model and recent extensions of that model by Petrusic and colleagues (Holyoak, 1978; Holyoak & Mah, 1982; Jamieson & Petrusic, 1975; Petrusic, 1992; Petrusic & Baranski, 1989a, 1989b), or propositional representations, such as Banks's discrete code or semantic coding model and recent updates of that model by Cech, Shoben, and their colleagues (Banks, 1977; Cech, 1995; Cech, Shoben, & Love, 1990; Shoben, Cech, Schwanenflugel, & Sailor, 1989; Shoben & Wilson, 1998). Although there is no generally accepted explanation of the semantic congruity effect in adults, the most current reformulation of the reference point model appears to provide the most unified account of a wide range of comparative judgment phenomena, including the semantic congruity effect (Petrusic, 1992).

Children's Acquisition of Dimensional Adjectives

Preschool children acquire dimensional adjectives slowly, not showing a full understanding of the meanings of common words until as late as 5 or 6 years of age (see Donaldson & Wales, 1970; Ehri, 1976; Klahr & Siegler, 1978; Maratsos, 1973; Sinclair-de-Zwart, 1969; Trabasso, 1977). During this period of acquisition they display two phenomena that bear a strong resemblance to the semantic congruity effect found in adult comparative judgments.

One phenomenon found in children's early errorful understanding of dimensional adjectives is referred to in this study as an *endpoint effect*. The endpoint effect is the finding that, when asked to label a series of objects in a categorical fashion (e.g., as big or little), young children consistently label only the extreme values of a series. For example, given a series of objects varying in size, young children typically label only the largest items as big and only the smallest ones as little while maintaining that the objects in between are neither big nor little (Ehri, 1976; Sera & Smith, 1987). With age, children progress from labeling only the endpoints to a categorical partitioning of the objects into two groups (Sera & Smith, 1987).

The second phenomenon is the finding that young children show robust semantic congruity effects in the accuracy of their comparative judgments during this period of acquisition. That is, a long history of results indicates that children have considerable difficulty making comparative judgments when the positions of the stimuli in the series are incongruent with the requested comparison (Clark, 1970; Ebeling & Gelman, 1988; Ehri, 1976; Ryalls, Winslow, & Smith, 1998; Sera & Smith, 1987; Smith, Ratterman, & Sera, 1988). For example, in one series of experiments, Smith and colleagues investigated children's understanding of the dimension words higher and lower (Smith et al., 1988). In these experiments children were presented with pairs of objects placed at various heights and asked to make comparative judgments (e.g., "Which is higher?" or "Which is lower?"). When presented with the highest pair of objects, young children correctly chose the higher one. When presented with the lowest pair of objects, young children correctly chose the lower one. However, when presented with mid-height objects, objects consistently classified as neither high nor low, children performed poorly both when asked to choose the higher one and when asked to choose the lower one. Most important, however, children's performance was extremely poor when they were asked to choose the lower of the two highest objects or the higher of the two lowest objects. Thus, it appears that what adults do more slowly, children have difficulty doing at all. Critically, endpoint and semantic congruity effects in children's performance are *context specific*, just as they are in adult performance (Ryalls et al., 1998; Smith, Cooney, & McCord, 1986), and are specific to an understanding of dimensional terms and not to all forms of comparison (Ryalls et al., 1998).

The Present Research

This review of the adult and developmental literatures is important for two reasons. First, it is important because any theory that attempts to explain the semantic congruity effect in adults should, at least, be compatible with the developmental trends and, at best, provide a unified account of the developmental and adult findings. Thus far, no attempt has been made by proponents of any theory to incorporate the developmental findings. Second, the joint comparison of these two literatures suggests that examining the acquisition of dimensional words may inform our understanding of the mature usage of these terms. For example, Ryalls et al. (1998) examined children's acquisition of the words higher and lower and found patterns of performance incompatible with most existing explanations of the semantic congruity effect. In addition, researchers have begun examining children's acquisition of novel dimensional words. The results of Ryalls (2000) indicate that the presence or absence of the semantic congruity effect in children's accuracy depends on the particular characteristics of learning. These findings raise the possibility that the presence or absence of the semantic congruity effect in the performance of adults taught new dimensional words might also depend on the particular characteristics of training, which may in turn constrain theories attempting to explain adult performance with conventional dimensional words.

Thus, in the present article we report the results of four experiments aimed at exploring adult acquisition of novel dimensional adjectives and at exploring the aspects of training that affect learning and that may or may not lead to semantic congruity effects in both accuracy and reaction time, utilizing both quantitative and qualitative analyses. Our working hypothesis was that the semantic congruity effect in adults is the developmental product of the way children acquire dimensional adjectives. From this perspective, we examined the acquisition of conventional dimensional adjectives and identified aspects that we believed might drive the formation of categories in labeling and create the semantic congruity effect in comparative judgments. We also asked whether category formation in learning is somehow critical to the emergence of the semantic congruity effect, as suggested by work with children (Ryalls, 2000). We modeled our adult learning task to be like children's learning tasks and manipulated aspects of training we hypothesized to be critical.

In our learning task we defined a series of seven objects that we treated as a linear order. We will call these seven items A, B, C, D, E, F, and G. Participants were taught the names for the two directions of difference along this series, the A-most direction and the G-most direction. The novel words used to describe the two directions, *mal* and *borg*, functioned in our task the same way conventional adjectives such as big and little function. That is, they were used to label individual objects and to label directions of difference in the series. In the experiments reported here, participants never saw the series as a whole. Instead, we taught people the names for the two directions in the following manner: We showed participants pairs of stimuli and labeled only one, for example, "This is *mal*" (or "This is *borg*"). In all phases, objects were labeled in a manner consistent with the defined series. That is, if A was the *mal* endpoint and G was the *borg* endpoint, then the A-most member of the pair was labeled as *mal* or the G-most member was labeled *borg*. Participants, however, were never told that the relation between the two items presented in a trial determined the labeling. This may seem to be a difficult task because the relational meanings of *mal* and *borg* were not explicit. However, although the relational nature of the meanings of *mal* and *borg* (or big and little) might not be immediately obvious from the labeling of a single member of a pair according to its relation to the other member in the pair, there is in fact a great deal of information in such input: As long as all pairwise comparisons are made, there is enough information from the pairwise comparison of objects to construct the series itself. In addition, recent evidence indicates that this type of labeling is commonly used by parents interacting with their young children (Ryalls, Hall, Johnson, Reimer, & Anthis, 1999).

Two aspects of the learning task seemed potentially important and were therefore manipulated. The first was the frequency with which individual objects in the stimulus series were labeled with each term. In the world, objects are not labeled with opposing dimensional adjectives (such as short and tall) equally often. Professional basketball players are much more likely to be called tall, whereas jockeys are more likely to be called short. Our question was whether this unequal or ordered labeling affects the ease of acquisition of novel dimensional adjectives and if it is required for the emergence of categories and a semantic congruity effect. This seemed possible because an endpoint could, as a product of being most often labeled by one term, become the "best exemplar" for a term, and thus a "reference point" of sorts. To investigate this possibility, we used two labeling conditions in Experiment 1. The first condition employed *ordered labeling* and thus mirrored the kind of labeling found using conventional, real-world terms; items closer to the *mal* end were labeled *mal* more often than *borg* and items closer to the *borg* end were labeled *borg* more often than *mal*. In the second condition, the *equal-labeling* condition, this information source was taken away. All objects explicitly labeled in training in the equal-labeling condition were labeled *mal* and *borg* an equal number of times.

The second potentially critical aspect of learning involved the nature of the stimulus continuum. Dimensional adjectives refer to quantitative dimensions, but these may be perceptual, such as size, or more abstract conceptual ones, such as intelligence. The semantic congruity effect in adults has been found with both types of dimensions, although the effect is more robust with symbolic or conceptual comparisons than with perceptual comparisons (see Petrusic & Baranski, 1989b). In Experiment 1 we asked if a perceptually (or conceptually) ordered series is necessary for the acquisition of novel dimensional adjectives and the development of a semantic congruity effect by varying the nature of the stimulus continuum. One group of participants learned *mal* and *borg* as they applied to the perceptually ordered cartoon stimuli in Figure 1. Another group learned *mal* and *borg* as they applied to the arbitrarily ordered shape stimuli, also illustrated in Figure 1. These particular shapes were chosen because they have no obvious natural ordering. That is, the perceptual information present in the shapes does not vary systematically as it does in the other naturally ordered series.[1]

The resulting experiment, Experiment 1, consisted of the four between-subjects conditions that derive from the combination of two factors: (a) perceptually ordered versus arbitrarily ordered stimuli and (b) ordered labeling versus equal labeling. These four conditions differed systematically in the amount and

[1]The arbitrarily ordered series of shapes was chosen over a random ordering of the cartoon figures because we believed that the perceptual information in the cartoon series would be too salient for the participants to ignore and thus preclude learning. Although this assumption remains to be tested, we believe that the difficulty that participants who trained with the arbitrarily ordered shapes had in learning and their anecdotal comments concerning their learning strategies justify our choice of stimuli.

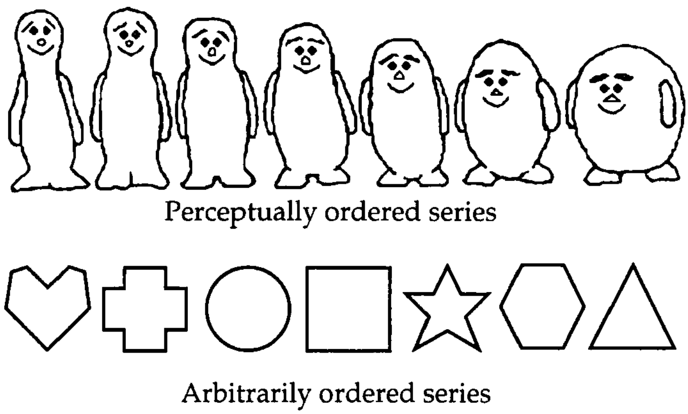

FIGURE 1. Perceptually ordered and arbitrarily ordered stimulus series used in Experiment 1.

kinds of information available to the learner about the ordering of objects in the series and thus about the opposing directions of comparison to which the words *mal* and *borg* referred. The condition that provided the most information—both perceptual information and frequency of labeling information—was very much like a natural dimension. The condition that provided the least amount of information—neither perceptual information nor frequency of labeling information—was very much unlike a natural language dimension. Thus, by comparing these four conditions we can ascertain how each affects the acquisition of novel dimensional adjectives and which of these sources of information, if any, are necessary for the emergence of a semantic congruity effect in comparative judgments.

In training participants we did not explicitly teach them to establish categories. However, throughout the course of learning we tested whether the participants had spontaneously developed categories.The role of categorization in the process of comparative judgment is unclear (see Shoben & Wilson, 1998); therefore, examining the spontaneous formation of categories may help us to understand how the two processes are related. In this experiment we also examined two comparative judgment measures—accuracy and reaction time. By collecting both accuracy and reaction time data we can examine the relationship between the semantic congruity effects in these two measures. As reviewed earlier, the semantic congruity effect in accuracy is routinely observed in children's performance. The semantic congruity effect in reaction time is extremely robust in adult performance. However, a semantic congruity effect in the accuracy of adult performance has been reported only once, despite the fact that there is reason to believe an effect should be observed in adults' accuracy (see Petrusic, 1992). In the present experiment we asked if a semantic congruity effect in accuracy is an obligatory part of the acquisition process by looking for the effect in adult acquisition.

EXPERIMENT 1

Method

Participants and Design

Forty undergraduates, ranging from 19 to 26 years of age, were recruited from a midwestern U.S. university and were given credit as part of a class requirement or received monetary compensation for participation. Each participant was randomly assigned to one of two stimulus conditions (perceptually ordered stimuli or arbitrarily ordered stimuli) and one of two labeling conditions (ordered labeling or equal labeling), resulting in a 2 × 2 factorial design. There were equal numbers of men and women in each unique condition. No other demographic information was collected.

Materials and Apparatus

Two stimulus continuums were employed in this experiment, a perceptually ordered stimulus series and an arbitrarily ordered stimulus series. In the perceptually ordered stimulus condition, the stimulus continuum consisted of seven cartoon figures (A–G) that varied systematically on a number of correlated dimensions (e.g., height, width, length of arms, bushiness of eyebrows, etc.; see Figure 1). Labeling of the two directions in the series was counterbalanced. That is, for half of the participants, stimulus A was the *mal*-est stimulus and G was the *borg*-est. For the other 10 participants labeling was reversed. Prior to analyses, half the data were reversed. Stimuli in the arbitrarily ordered condition consisted of seven geometric shapes (see Figure 1). The order of the shapes in this series was randomly determined. Ten participants were assigned to each of two different randomly determined shape orders.

Stimuli used in training and all nonspeeded tests were printed on 4 × 5 in. sheets of paper. Stimuli used in the speeded comparative-judgment task were printed on 2 × 3 in. cards. The speeded comparative-judgment task was conducted using a tachistoscope connected to a timer.

Procedure

Training and testing consisted of several phases, each of which is described as follows.

Training. Participants were taught two new words that functioned as dimensional adjectives in a pairwise-learning procedure. The participants were instructed to learn what the words *mal* and *borg* meant as they applied to the pictures presented. Training was conducted in blocks of 50 trials. During training, participants were presented with pairs of stimuli drawn from the seven-item series. Each pair was presented individually, one item at a time, with the second object in each pair labeled as either *mal* or *borg* using the sentence "This one is X." Half of the participants were assigned to the ordered-labeling condition, and half were assigned to the equal-labeling condition.

In the ordered-labeling condition, pairs of stimuli were constructed so that the closer a stimulus was to one end of the continuum, the more likely it was to be labeled with the term corresponding to that endpoint. For example, in the condition in which A was the *mal*-est stimulus, it was labeled *mal* 100% of the time and never labeled *borg*. B was the *mal* member of the pair 88% of the time and the *borg* member 12% of the time (when paired with stimulus A); C was labeled *mal* 63% of the time and *borg* 27% of the time; D was labeled *mal* and *borg* equally often. From this point, the pattern was reversed until stimulus F was labeled as *borg* 100% of the time. Each of the seven items appeared an approximately equal number of times per block (range = 14–15). Every possible pair was presented 2–3 times per block and was presented an equal number of times across blocks. The 50 stimulus pairs in each training block were presented in a random order. An example of one of the ordered-labeling training blocks is presented in the Appendix.

In the equal-labeling condition, there was no ordering information available on the frequency with which each stimulus was labeled with each term. The training pairs were constructed in such a way that, although every possible pair was presented, labeling was such that each item, excluding the endpoints, was explicitly labeled *mal* and *borg* an equal number of times (5 times as *mal* and 5 times as *borg*). In this condition the endpoints (Stimuli A and G) were never explicitly labeled. Each stimulus item appeared an approximately equal number of times in each training block (range = 14–15), and each possible pair appeared an equal number of times across blocks. An example of one of the training blocks involving equal-labeling is presented in the Appendix.

Categorization test. In the categorization test, participants were presented with each of the seven stimulus items individually and were asked "Is this *mal*?" (or on other trials, "Is this *borg*?"). Participants responded by circling yes or no on a response sheet. Each of the seven items was randomly presented four times for each of the two terms (*mal* and *borg*) for a total of 56 trials per test. No feedback was provided in the categorization test.

Comparative learning test. In this nonspeeded comparative-judgment test, participants were presented with the six pairs of consecutive stimuli[2] (i.e., A and B,

[2]It should be noted that in training, participants were presented with all possible pairs. Testing with only the consecutive pairs in the comparative judgment test precludes examination of the data for distance effects.

B and C, C and D, etc.) and were asked, "Which is *mal*?" (or on other trials "Which is *borg*?"). The pairs of stimuli were presented together on the page. Participants responded by circling left or right, corresponding to the side of the page of the correct response, on a response sheet. The side of the correct response was counterbalanced. Each of the six pairs was randomly presented two times for each of the two terms (*mal* and *borg*) for a total of 24 trials per test. No feedback was provided in the comparative learning test.

Speeded comparative-judgment test. This test was similar to the nonspeeded test of learning in that pairs of stimuli were presented and participants were asked to indicate which member of the pair was *mal* or which was *borg*. This task differed in that stimuli were presented via a tachistoscope and participants' responses were timed. Participants responded by pushing a left button or a right button corresponding to the side of the correct response. Side of the correct choice was counterbalanced. Participants were instructed to respond as quickly as possible while maintaining high accuracy. Participants in Experiment 1 were tested on only the endpoint pairs (AB and FG). Responses were recorded on audiotape and later transcribed. Each of the two pairs was presented 40 times for each of the two terms (*mal* and *borg*) for a total of 160 randomly ordered trials. During the experiment, participants were given the opportunity to take a short break after every 80 trials. No feedback was provided in this test.

The experiment progressed as follows: In Session 1 participants were first pretested with the categorization test to verify that there was no pre-existing tendency to categorize the stimuli with the words *mal* and *borg*. There was no such pre-existing bias, so these data will not be reported. The pretest was followed by two training/testing blocks (Blocks 1 and 2), each of which consisted of one training block, one comparative learning test, and one categorization test. In Session 2, participants completed two more training/testing blocks (Blocks 3 and 4). In Session 3, participants were tested in the speeded comparative-judgment test. The three sessions lasted approximately 1 hr each and were conducted on 3 consecutive days.

Results and Discussion

We first examined how well participants learned the comparative meanings of the terms in the four conditions. The percentages of correct responses in the comparative learning test on the last trial block of the last session (Session 2) were analyzed in an analysis of variance (ANOVA) for a 4 (condition) × 6 (stimulus pair) × 2 (term) mixed design. Results indicated a main effect of condition, $F(3, 36) = 5.83$, $p < .01$. Post hoc analyses indicated that participants in the perceptually ordered stimuli/ordered-labeling condition (with an overall mean of 84% correct) performed better than participants in the perceptually ordered stimuli/equal-labeling, arbitrarily ordered stimuli/ordered-labeling, and arbitrarily ordered stimuli/equal-labeling conditions (with overall means, respectively, of

61%, 49%, and 55% correct), which did not differ. Chance performance in this task was 50%.

Despite the low performance overall in some of the conditions, individual participants' learning was highly variable within and between conditions. To examine individual differences in learning, "learners" were defined as those participants whose performance reached the criterion of 70% correct overall on the last trial block of Session 2. Eighteen participants performed at a level of 71% correct or above. The 70% cutoff was chosen in this study because of a natural division in participant performance. Nine out of 10 people in the perceptually ordered stimuli/ordered-labeling condition were learners by this criterion; 5 out of 10 people in the perceptually ordered stimuli/equal-labeling condition were learners; 3 out of 10 people in the arbitrarily ordered stimuli/ordered-labeling condition were learners; and finally, only 1 out of 10 people in the arbitrarily ordered stimuli/equal-labeling condition learned by this criterion.

In the next analyses we examined, for the learners in each condition, the relation between the development of an understanding of comparative meaning, the categorization of individual objects, and the presence of a semantic congruity effect in speeded comparative judgments.

Perceptually Ordered Stimuli/Ordered-Labeling Condition

Comparative learning test. The perceptually ordered stimuli/ordered-labeling condition was the most natural and presented participants with the most sources of information about the series and with the directional meaning of the words: perceptual information, ordered-labeling information, and pairwise comparisons. Thus, it is not surprising that nearly all of the participants learned in this condition. We first asked whether these participants showed the kind of congruity effects common to young children's comparative judgments during the course of learning. That is, we asked if there was a semantic congruity effect in accuracy.

We addressed this question by examining changes in the accuracy of comparative judgments as a function of the position of the pair in the series during learning. These data are shown in Figure 2. An ANOVA for a 4 (trial block) × 6 (stimulus pair) × 2 (term) within-subject design revealed a main effect of stimulus pair, $F(5, 45) = 4.97$, $p < .01$, and a significant three-way interaction among trial block, stimulus pair, and term, $F(15, 135) = 2.12$, $p < .01$. As can be seen in Figure 2, participants did better on the endpoints overall, resulting in the main effect of stimulus pair. The three-way interaction resulted because participants showed a semantic congruity effect in the accuracy of their comparative judgments in the earliest stage of learning (Trial Block 1) but not in later blocks. To confirm this conclusion, an ANOVA for a 6 (stimulus pair) × 2 (term) within-subject design on Trial Block 1 alone indicated a significant interaction between stimulus pair and term, $F(5, 45) = 2.96$, $p < .05$. That is, early in learning, people were better at making *mal* judgments than *borg* judgments near the *mal* end of the series and were better at making *borg* judgments than *mal* judgments near the *borg* end of the series. This effect in Trial Block 1 is only the second semantic congruity effect in accuracy ever reported for adults (see Petrusic, 1992, for the first). Analyses on the other trial blocks indicated no stimulus pair by term interactions. The fact that this effect was not evident in later blocks may reflect the lack of pressure to respond rapidly, or the participants' realization that the words referred to opposite directions (*mal* means not *borg*), or both. That is, in

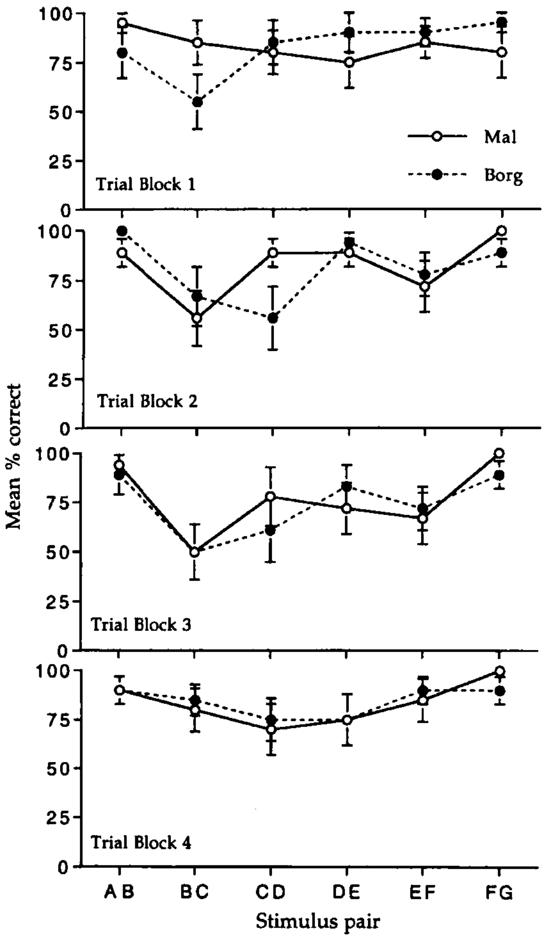

FIGURE 2. Comparative learning test results in the perceptually ordered stimuli/ordered-labeling condition of Experiment 1, by stimulus pair, term, and trial block. *n* = 10. Vertical lines depict standard error (*SE*).

the unspeeded learning task, participants may not show a semantic congruity effect (despite a perhaps underlying one) once they realize that, in a pair, the *mal* one is not the *borg* one. To the degree that participants know the *mal* item in any pair, they will be able to respond *borg* to the not-*mal* one.

Categorization test. Like children learning real dimension words, adults in this condition also spontaneously formed categories. Figure 3 shows the frequency with

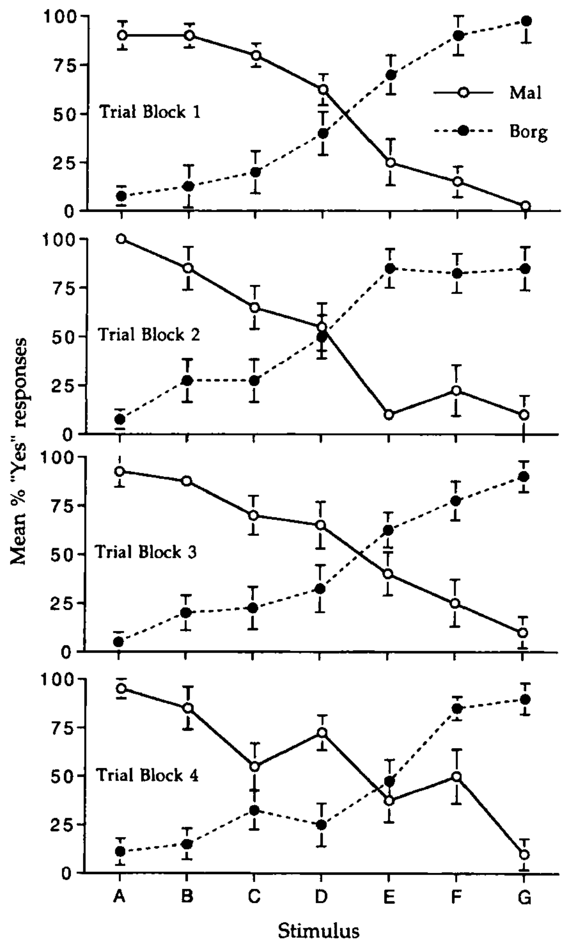

FIGURE 3. Categorization test results in the perceptually ordered stimuli/ordered-labeling condition of Experiment 1, by stimulus, term, and trial block. *n* = 10. Vertical lines depict *SE*.

which participants labeled each item in the series as *mal* and *borg*. Yes responses to the questions "Is this one *mal*?" and "Is this one *borg*?" in the perceptually ordered stimuli/ordered-labeling condition were submitted to an ANOVA for a 4 (trial block) × 7 (stimulus) × 2 (term) within-subject design. Results revealed a main effect of stimulus, $F(6, 54) = 2.41$, $p < .05$, and a significant two-way interaction between stimulus and term, $F(6, 54) = 29.31$, $p < .01$. The lack of a main effect of trial block indicates that participants' categorization of the stimuli did not change with training. The finding of strong categorization of the individual stimuli in the series is notable because participants were trained only on the comparative meanings of the words and were never given feedback in the categorization test.

In addition to the group analyses, each participant's performance in the categorization test was classified qualitatively according to the degree of catego-

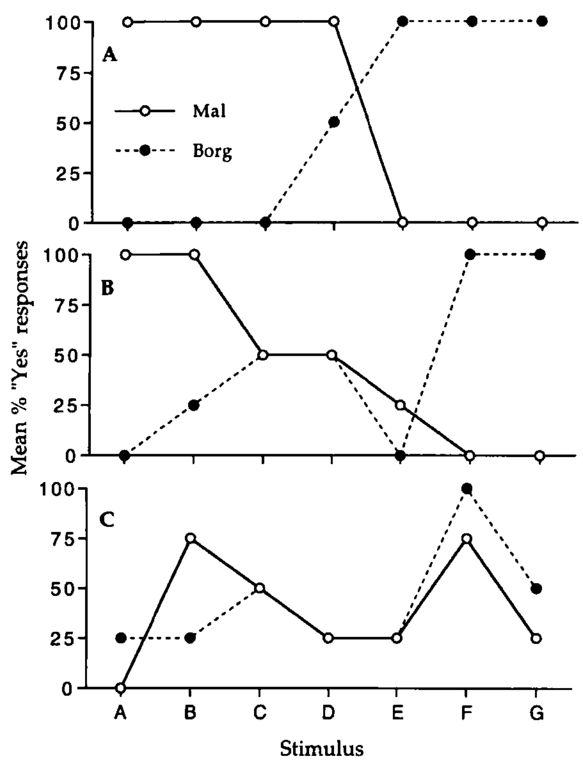

FIGURE 4. Examples of categorization test data are classified as follows: (A) full category formation, (B) endpoint category formation, and (C) other category formation.

TABLE 1
Number of Participants in Each Condition Classified in Each Categorization Classification in Experiment 1

	Classification of categorization performance		
Condition	Categories	Endpoints	Other
Perceptually ordered stimuli/ordered labeling	9	1	0
Perceptually ordered stimuli/equal labeling	2	5	3
Arbitrarily ordered stimuli/ordered labeling	7	2	1
Arbitrarily ordered stimuli/equal labeling	0	1	9

rization of the seven stimulus items. For our purposes, an individual stimulus item was said to be "categorized" if that item was classified into one category 75% of the time or more and classified into the other category less than 25% of the time. Each participant's performance was designated as fitting one of three types of categorization behavior, defined as follows: *Full categorization* was defined as categorization of six of the seven stimulus items, including the endpoints (see Figure 4A). *Endpoint categorization* was defined as appropriate categorization of the endpoints but less than full categorization (Figure 4B). *Other* was defined as any other organization or lack thereof (Figure 4C). Participants grouped in this class sometimes showed no consistent pattern of categorization at all. Others tended to categorize either the middle five or all of the stimuli as both *mal* and *borg*. This pattern makes some sense if one is attending only to the input and not constructing a linear representation of the series because all stimuli, except the endpoints, are labeled as both *mal* and *borg* at least once during training.

The number of participants whose categorization performance fit each of these definitions in each condition is shown in Table 1. In the perceptually ordered stimuli/ordered-labeling condition, 9 out of 10 participants formed full categories and 1 individual demonstrated endpoint categorization at some point in training. This last participant was not the same individual classified as a nonlearner in the comparative learning test.

Speeded comparative-judgment task. Accuracy on the speeded comparative-judgment task in this condition was high, with an overall mean of 89% correct. Average response times for all correct responses were submitted to an ANOVA for a 2 (stimulus pair: AB and FG) × 2 (term) within-subject design. Results revealed a significant two-way interaction between stimulus pair and term, $F(1, 8) = 14.55$, $p < .01$. This is a significant semantic congruity effect (see Table 2). That is, participants were faster to choose the *mal*-est stimulus A as the *mal* member of the pair AB than to choose B as the *borg* member. Conversely, they were faster to

TABLE 2
Mean Reaction Time (in ms) and Standard Errors for Learners' Speeded Comparative Judgments in Experiment 1, by Condition, Stimulus Pair, and Word

	Stimulus pair			
	AB		FG	
Condition/word	*M*	*SE*	*M*	*SE*
Perceptually ordered stimuli/ordered labeling ($n = 10$)				
Mal	536	50	581	51
Borg	576	58	509	41
Perceptually ordered stimuli/equal labeling ($n = 5$)				
Mal	580	142	670	202
Borg	618	133	557	142
Arbitrarily ordered stimuli/ordered labeling ($n = 3$)				
Mal	431	15	600	43
Borg	517	37	571	36

choose the *borg*-est stimulus G as the *borg* member of the pair FG than to choose F as the *mal* member.

Summary. Adults in this condition passed through the same stages as young children who are first learning dimensional words, showing a semantic congruity effect in the accuracy of their comparative judgments early in learning, partitioning the stimulus series into two discrete categories in their categorization behavior, and displaying a semantic congruity effect in their speeded comparative judgments at the end of training.

Perceptually Ordered Stimuli/Equal-Labeling Condition

Comparative learning test. This condition presented the participants with a perceptually ordered continuum but no information on the frequency with which individual objects were labeled with the two terms. Performance in the comparative learning test indicated that 5 participants met the criterion for learners, with an average overall performance of 89% correct on the last trial block of Session 2. For the 5 remaining nonlearners, overall performance rate was 23% correct on the last trial block of Session 2. The learners' data were submitted to an ANOVA for a 4 (trial block) × 6 (stimulus pair) × 2 (term) within-subject design, which revealed no significant main effects or interactions. A similar analysis of the nonlearners' data also indicated no main effects or interactions. Importantly, at no point in learning was there an advantage for comparisons involving a named endpoint. That is, there was no semantic congruity effect in accuracy for either the learners or nonlearners.

Categorization test. Figure 5 demonstrates the frequency with which the learners labeled each item in the series as *mal* and *borg*. The participants who learned to make comparative judgments also exhibited a tendency toward categorization of the stimuli. Specifically, over all trial blocks, learners strongly categorized stimuli A and B as *mal* and stimulus G as *borg*. The learners' yes responses were submitted to an ANOVA for a 4 (trial block) × 7 (stimulus) × 2 (term) within-subject design. Results revealed a significant two-way interaction between stimulus and term, $F(6, 24) = 7.05$, $p < .01$. In terms of individual performance, among the 5

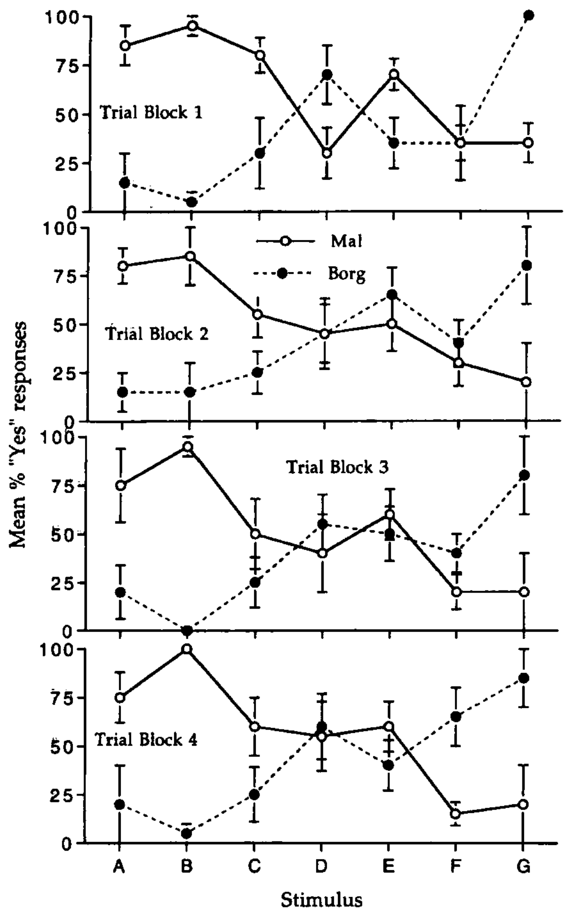

FIGURE 5. Categorization test results for learners in the perceptually ordered stimuli/equal-labeling condition of Experiment 1, by stimulus, term, and trial block. *n* = 5. Vertical lines depict *SE*.

learners, 1 participant formed full categories and 4 demonstrated endpoint categorization. This was not the case for the nonlearners. A similar ANOVA performed on the 5 nonlearners' categorization data indicated no significant main effects or interactions although, as is evident in Table 1, examination of each nonlearner's data indicated that 1 exhibited full categorization and 1 exhibited endpoint categorization.

Thus, in this condition we see a very different pattern of results, as compared with the perceptually ordered stimuli/ordered-labeling condition. Only half of the participants learned to make accurate comparative judgments, and participants were less likely to show strong categorization in labeling the stimuli, although learners were more likely to show categorization than nonlearners. Moreover, at no point did the 5 learners show a semantic congruity effect in the accuracy of their comparative judgments. This lack of a semantic congruity effect in learning might be due to the slowness of learning. If participants learned the relative meaning of each term after they realized that in every pair there is one *mal* object and one *borg* object, then they would not show a semantic congruity effect in learning but rather comparable performance on the two opposing terms at all positions in the series. If a semantic congruity effect in accuracy is critical to one in speeded judgments, then the learners should not show a semantic congruity effect in their speeded comparative judgments. If, however, forming categories in which (at least) the endpoints of the series are strongly placed in one category or the other is related to a semantic congruity effect in reaction time, then we might expect the learners to show such a semantic congruity effect.

Speeded comparative-judgment task. Learners' reaction times are displayed in Table 2. The mean error rate for the learners in this task was 7%. The learners' average response times for all correct responses were submitted to an ANOVA for a 2 (stimulus pair) × 2 (term) within-subject design. This analysis revealed a significant pair by term interaction, $F(1, 4) = 8.82$, $p < .05$. This is a significant semantic congruity effect, similar to that found in the perceptually ordered stimuli/ordered-labeling condition. The nonlearners' performance was not assessed as a group because the mean error rate for the nonlearners in this task was 50%. However, we did examine the performance of the 2 participants who showed some organization of the series in the categorization test—that is, full and endpoint categorization. Specifically, we submitted all correct reaction times for each participant to individual analyses. All 5 learners exhibited significant stimulus pair by term interactions—that is, significant semantic congruity effects. The 1 nonlearner who displayed endpoint categorization in labeling also showed a reliable semantic congruity effect. The nonlearner who demonstrated full categorization, however, did not show a semantic congruity effect.

Summary. In the perceptually ordered stimuli/equal-labeling condition, pairwise comparison information and perceptual information were available to help participants learn the meanings of the terms. However, the information from the differential labeling of items in the series by the two terms was not available. Half of the participants learned to make comparative judgments in this condition, compared with 9 out of 10 of the participants in the perceptually ordered stimuli/ordered-labeling condition. This finding suggests that participants made use of the ordered linguistic input, namely, the frequency with which individual objects were labeled *mal* and *borg*, in the first condition. However, pairwise comparison information and a perceptual continuum were sufficient for 7 out of 10 participants to demonstrate categorization of the stimuli and for 6 out of 10 to show a semantic congruity effect in reaction time at the end of training.

Arbitrarily Ordered Stimuli/Ordered-Labeling Condition

Comparative learning test. In the next condition, the arbitrarily ordered stimuli/ordered-labeling condition, there was no perceptual information in the stimulus series itself concerning the appropriate order. There was, however, ordering information in the frequency of labeling of the individual stimuli with each term in pairwise comparison. This amount of information was sufficient for only 3 out of the 10 participants to meet the criterion defined for learners. The 3 participants clearly did learn the relative meaning of *mal* and *borg*, performing at a level of 81% correct on the last trial block of the comparative learning test. The 7 nonlearners also clearly did not learn, performing at a level of 26% correct on the last trial block of the comparative learning test. The learners' data were submitted to an ANOVA for a 4 (trial block) × 6 (stimulus pair) × 2 (term) within-subject design, which revealed a significant effect of trial block, $F(3, 6) = 4.90$, $p < .05$, and a significant interaction between trial block and term, $F(3, 6) = 12.00$, $p < .01$. The main effect of trial block indicates that performance improved across trial blocks. The significant trial block by term interaction resulted because, on Trial Block 1, performance on *mal* judgments exceeded performance on *borg* judgments, but by Trial Block 4 this order was reversed. Again, we saw no evidence of a semantic congruity effect in the accuracy of performance on the part of the learners at any point in learning. Analysis of the nonlearners' data indicated no main effects or interactions, and there was no evidence in the nonlearners' data of a semantic congruity effect at any point.

Categorization test. The learners' labeling of the stimuli became more categorical across trial blocks (see Figure 6). Learners' yes responses were submitted to an ANOVA for a 4 (trial block) × 7 (stimulus) × 2 (term) within-subject design. Results revealed a two-way interaction for stimulus and term, $F(6, 12) = 4.70$, $p < .01$, and a three-way interaction for trial block, stimulus, and term, $F(18, 36) = 2.47$, $p < .05$. Individually, the 3 learners all partitioned the stimulus series into full categories at some point in learning. The nonlearners' organization also became more categorical across trial blocks (Figure 7). An analysis of the 7 nonlearners' categorization data indicated significant two-way interactions between trial block and stimulus, $F(18, 108) = 1.97, p < .05$, and stimulus and term, $F(6, 36) = 8.36, p < .01$. Four of the 7 nonlearners formed full categories, and 2 demonstrated end-point categorization. Thus, in this condition few participants learned the relative meaning of *mal* and *borg*. However, overall, participants did categorize the stimuli.

Speeded comparative-judgment task. The results in the speeded comparative-judgment task for the 3 learners are shown in Table 2. Overall, these participants per-

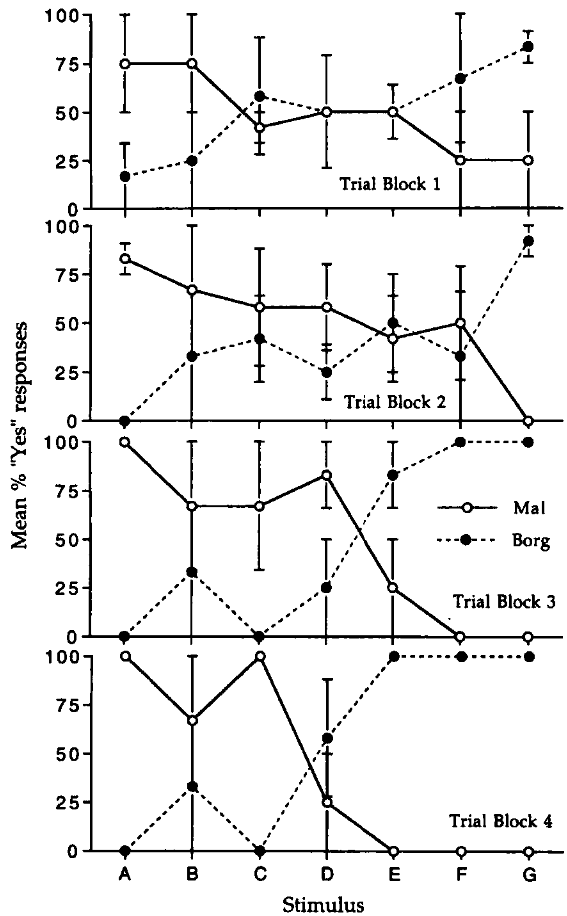

FIGURE 6. Categorization test results for learners in the arbitrarily ordered stimuli/ordered-labeling condition of Experiment 1, by stimulus, term, and trial block. *n* = 3. Vertical lines depict *SE*.

formed at a level of 84% correct on this task. The learners' average response times for all correct responses were submitted to an ANOVA for a 2 (stimulus pair) × 2 (term) within-subject design. Although the interaction in the group data did not reach significance, analyses of each individual's data indicated that 2 of the 3 learners produced a significant semantic congruity effect. Overall, nonlearners performed at a level of 54%. Only 1 nonlearner performed with accuracy high enough to be considered for analysis (77% correct overall). Analysis of that participant's

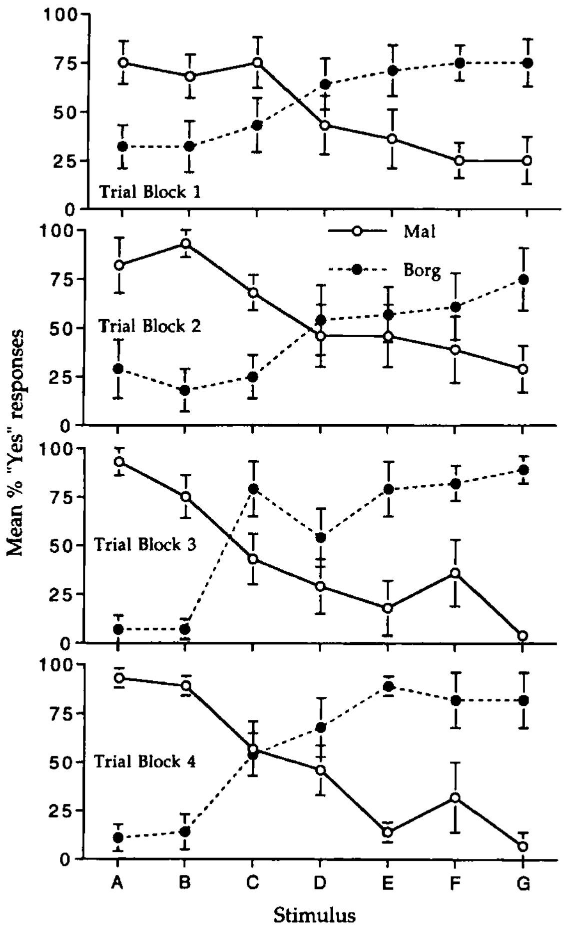

FIGURE 7. Categorization test results for nonlearners in the arbitrarily ordered stimuli/ordered-labeling condition of Experiment 1, by stimulus, term, and trial block. *n* = 7. Vertical lines depict *SE*.

data indicated a significant semantic congruity effect in reaction time. This individual had also consistently categorized the endpoints in the categorization task.

Summary. In this condition most of the participants did not learn the meaning of *mal* and *borg*. Those who learned, however, did demonstrate categorization of the stimuli in the categorization test, and 2 of the 3 learners showed clear evidence of a semantic congruity effect in their speeded comparative judgments. Although the 7 nonlearners did learn categories, they did not show a semantic congruity effect in their speeded comparative judgments.

Arbitrarily Ordered Stimuli/Equal-Labeling Condition

Comparative learning test. This condition was by far the most unnatural condition because there was no perceptual information available with which to order the series and no ordering information in the frequency of labeling with each term. The only information available came from the pairwise comparisons themselves. One participant reached the criterion to be classified as a learner with an overall average of 79% correct on the last trial block of Session 2. The remaining 9 nonlearners achieved an average of 52% correct overall on the same trial block. Because performance was so poor in this condition, no group analyses were conducted.

Categorization test. A single participant in this condition exhibited endpoint categorization. This was not the same participant identified as a learner in the comparative learning test.

Speeded comparative-judgment task. Nine of the 10 participants in this condition performed so poorly as to preclude group analysis. A single participant responded accurately enough for analyses but showed only a main effect of stimulus pair and no semantic congruity effect. This participant was neither the one identified as a learner in the comparative learning test nor the one displaying endpoint categories in the categorization task.

General Summary

In summary, the results of Experiment 1 tell us about the acquisition of relational words, the formation of categories, and the emergence of the semantic congruity effect. With regard to acquisition of relational words by adults, performance on the comparative learning test depended on the amount and the kind of information available to the participants. Those participants with the most sources of information available to them performed the best overall and showed evidence of a semantic congruity effect early in learning. Those participants with the fewest sources of information available to them demonstrated almost no learning. For the two conditions with intermediate levels of information available, it appears that the perceptual information available in stimulus series in the perceptually ordered stimuli/equal-labeling condition was more beneficial than the frequency of labeling information available in the arbitrarily ordered stimuli/ordered-labeling condition. More participants demonstrated learning in the perceptually ordered stimuli/equal-labeling condition than in the arbitrarily ordered stimuli/ordered-labeling condition, and perhaps they did so earlier in training.

The results were similar with regard to categorization. Participants with the most sources of information available were more likely to produce categories. In the perceptually ordered stimuli/ordered-labeling condition, the condition with the most information available, all participants produced strong categories. In the condition with the least amount of information available, the arbitrarily ordered stimuli/equal-labeling condition, only a single participant demonstrated endpoint categorization. With regard to the two conditions with an intermediate level of information, the pattern was reversed from the comparative learning test. Whereas in the comparative learning test more participants in the perceptually ordered stimuli/equal-labeling condition demonstrated learning than in the arbitrarily ordered stimuli/ordered-labeling condition, the pattern in categorization was opposite. Slightly more participants in the arbitrarily ordered stimuli/ordered-labeling condition formed categories than did those in the perceptually ordered stimuli/equal condition (9 vs. 7).

With regard to the emergence of the semantic congruity effect, the conclusions are different for accuracy and reaction time measures. That is, although the semantic congruity effect is quite robust in children's accuracy, the present data indicate that a semantic congruity effect is by no means obligatory in accuracy in adults' acquisition of dimensional adjectives. With respect to the relationship between semantic congruity effect in accuracy and reaction time, the results indicate that the effect need not be evident in errors in learning to show itself in reaction time. Instead, the results suggest that if adults learn to make the comparisons they will show a semantic congruity effect in reaction time at the end of learning and they will form categories.

These results also have implications for the relation between success in learning the words and the semantic congruity effect in reaction time and between categorization and the semantic congruity effect in reaction time. In this study, all three aspects of performance—learning the relational meaning, grouping items into categories, and exhibiting the semantic congruity effect in reaction time—went together. These results suggest that categorization may be a natural consequence of learning dimensional meanings regardless of how these meanings are learned. Consistent categorization alone is not sufficient for a semantic congruity effect; 9 participants across conditions formed categories but did not produce a semantic congruity effect in speeded comparative judgments. However, the results of this first experiment indicate that categorization may be necessary: All 18 of the participants who demonstrated a semantic congruity effect in their speeded comparative judgments had previously shown some tendency toward categorization. With regard to learning, nearly all of the participants who produced a semantic congruity effect in reaction time were learners ($n = 15$). Only 2 participants who failed to reach our learning criterion produced a semantic congruity effect in reaction time. Taken together, these results suggest that a semantic congruity effect in reaction time and category formation are the consequences of knowing the relational meaning of words regardless of how that meaning is learned.

Although the results of Experiment 1 are suggestive, there are a number of limitations. First, the number of learners in some of the conditions was very low. Thus, it is difficult to have confidence in all of the previous conclusions. Second, in the speeded comparative-judgment task, only the endpoint pairs were tested. This raises the possibility that the effect we obtained is only an "endpoint" effect and not a true semantic congruity effect (Potts, 1972). Finally, in Experiment 1 we chose to train participants using the unmarked form of the novel dimensional adjectives (*mal* and *borg*) instead of the marked forms (*maller* and *borger*). Our reasoning was that young children who are first learning dimensional words do not know that these words refer to dimensions, and, therefore, we wanted adults in our experiment to be similarly blind to the intended meaning. In addition, the marked form is used very rarely by mothers in conversation with their children (Ryalls et al., 1999). However, it is possible that our assumptions were incorrect and that we may have confused participants by not following the rules of adjectival semantics, which generally distinguish absolute and comparative meanings, thereby altering the pattern of acquisition.

Because of these concerns, we replicated two conditions of Experiment 1 in Experiments 2 and 3. Specifically, we replicated the two conditions of intermediate difficulty, the perceptually ordered stimuli/equal-labeling and the arbitrarily ordered stimuli/ordered-labeling conditions, altering particular aspects of the first experiment in order to validate our original findings and address remaining questions.

EXPERIMENT 2

In this experiment we replicated the perceptually ordered stimuli/equal-labeling condition from Experiment 1. This training condition provided perceptual information concerning the ordering of the series but no information in the relative frequency of labeling. As mentioned previously, a number of changes were made in the procedure used in this replication experiment. The biggest difference was that "word form" was manipulated in this experiment (and in Experiment 3). Participants were assigned to either the "unmarked" word-form condition or the "marked" word-form condition. The unmarked form was the same form used in all conditions of Experiment 1 in which comparative judgment questions were asked using the syntactically unmarked form of the dimension word (e.g., "Which one is tall?" or "Which one is *borg*?"). In contrast, the marked form involves explicitly marking dimension words as comparatives by adding the "-er" suffix when comparing two objects (e.g., "Which one is taller?" or "Which one is *borger*?").

In addition, we attempted to correct for the low number of learners in Experiment 1. In that experiment we trained 10 participants and then analyzed data from the participants who learned. In Experiment 2, the learning status of the participants was monitored and data collection continued until there were 10 learners in each word-form condition ($N = 20$).[3] The third major change was that reaction time data were collected for all contiguous pairs of stimuli (AB, BC, etc.) to ensure that the effects observed in Experiment 1 were true semantic congruity effects and not merely a result of the fact that the endpoints of our series were treated differently in training than the other stimuli. Finally, participants in this experiment completed more trial blocks, and accuracy and reaction time were measured concurrently throughout the procedure (instead of measuring reaction time only at the end). Because participants completed more blocks, we were able to examine potential individual differences in the ease of acquisition and the potential changes in performance with practice (among participants who learned early). A number of other minor changes were also necessary because this experiment was conducted using a computer instead of pencil and paper. These changes will be discussed as follows. Similar to Experiment 1, participants were classified qualitatively as learners or nonlearners and as categorizers or noncategorizers, and these qualitative distinctions were used as variables in some analyses.

Method

Participants

A total of 26 undergraduates ranging in age from 20 to 28 years were recruited from a midwestern U.S. university and were given extra credit in a psychology course for participation. As discussed earlier, the learning status of the participants was monitored and data collection continued until there was a total of 20 learners—10 learners in the unmarked word-form condition and 10 learners in the marked word-form condition. There were approximately equal numbers of male and female learners in each unique condition. The data from the 6 nonlearners were not analyzed in detail and will be discussed only in terms of the relative ease of acquisition in the different conditions.

[3]The ideal situation would be to take 10 participants and train them until they all learned. However, we chose the present compromise for two reasons. One, the majority of our participants were working for extra credit and they received credit for participating, regardless of how well they performed. Two, based on extensive pilot testing, we had reason to believe that some apparently motivated participants could not learn, especially in conditions involving the arbitrarily ordered series.

Materials and Apparatus

The perceptually ordered series used in this experiment was constructed by taking the cartoon characters used in Experiment 1, removing all lines and features within the outline of each shape, and filling the entire shape with gray shading. This process essentially turned the stimuli from animate characters to inanimate blobs but left several redundant cues concerning the appropriate ordering of the objects. Stimulus presentation and data collection were controlled by the application SuperLab (Cedrus, Inc.) running on a Power Macintosh 6100/60.

Procedure

The procedure for this experiment was similar to that of Experiment 1 except that it was conducted using a computer. Participants completed 15 blocks, with each block consisting of three parts: a training phase, a categorization test, and a comparative learning test. Each part is described as follows.

Training. Like Experiment 1, the training was conducted in blocks of 50 trials. During training, participants were presented with pairs of stimuli taken from the seven-item series. Each training trial consisted of two events. First, the question was presented on the screen for 350 ms. Participants in the unmarked word-form condition were asked either "Which is *mal*?" or "Which is *borg*?" Participants in the marked word-form condition were asked either "Which is *maller*?" or "Which is *borger*?" Then a stimulus pair appeared and remained on the screen until the participant chose a stimulus by pushing the *f* key for the object on the left or the *j* key for the object on the right. If the participant chose correctly, the stimulus pair disappeared and the next trial began. If the participant was incorrect, the stimulus pair remained on the screen until the correct choice was made. By attending to this feedback, participants could learn the meaning of the words. The 50 pairs were presented in random order. As discussed earlier, the frequency of labeling in this experiment was identical to the equal labeling used in Experiment 1 (see the Appendix for an example of 1 training block).

Categorization test. On each categorization trial, participants were first presented with a screen that asked "Is this *mal*?" (or on other trials, "Is this *borg*?"). The unmarked usage is the only form appropriate for this task and, thus, did not vary between word-form conditions. This screen was replaced after 350 ms by a screen containing one of the seven stimulus items. The stimulus remained on the screen until participants responded either yes or no by pushing either the *d* or *k* key, respectively. Participants were given no feedback. Each of the seven items was randomly presented one time for each of the two terms (*mal* and *borg*) for a total of 14 trials per test. This test was shorter than the test used in Experiment 1, in which 56 trials were presented in each of the four blocks.

Comparative learning test. The trials in this test were similar to the trials in the training phase except for the following differences. First, no feedback was given in this phase. The screen displaying the stimulus pair disappeared as soon as a key was pushed, regardless of whether it was the correct choice or not. Second, each of the six contiguous pairs (AB, BC, CD, etc.) was randomly presented two times for each of the two terms (*mal/maller* and *borg/borger*) for a total of 24 trials per test. This is the same number of trials as was presented in the same phase in Experiment 1. However, this is significantly fewer trials than the 160 that were included in the speeded comparative judgment test in Experiment 1. That phase was not required in this experiment because SuperLab records both accuracy and reaction time.

Participants were tested individually in a single session lasting 1.5–2 hr. All participants completed 15 blocks, each consisting of the three phases described previously. Participants were allowed to take breaks between blocks if they desired. Participants were randomly assigned to either the unmarked word-form condition or the marked word-form condition (henceforth referred to as the unmarked and marked conditions).

Results and Discussion

Ease of Acquisition

Two measures of the ease of acquisition are the number of participants required in each condition to reach 10 learners and, within the learners, the block in which the learning criterion was met and maintained. Each measure will be discussed in turn.

In this experiment, participants had to average 70% correct on the last 5 trial blocks of training in order to be considered learners. However, as will be discussed, there was a natural division in subject performance that made the issue of a learning criterion moot. Overall, slightly more than 3 out of 4 participants (77%) reached the learning criteria in the two word-form conditions, which did not differ. Each condition required the testing of 13 participants to obtain the desired 10 learners. The 20 participants identified as learners averaged 95% correct on the last 5 trial blocks (range = 81–100%). The 6 participants identified as nonlearners averaged 57% correct on the last 5 trial blocks (range = 42–60% correct). Chance performance in this task was 50%.

Although the number of participants necessary to reach 10 learners is one way to judge ease of acquisition, it is a fairly rough measure. Undergraduates participated in this study to receive course credit. They received credit regardless of their performance. The procedure in this experiment was long and tiresome. Thus, we do not know if participants failed to learn because they were unable to learn (in the allotted amount of time) or if they failed to learn simply because they did not try hard enough. Another approach to measuring ease of acquisition is to limit comparison to the learners, who were apparently sufficiently motivated, and compare ease of acquisition within the course of the experiment.

Examination of this measure indicates that learners in the present experiment needed an average of only 1.75 blocks to reach or exceed criterion, or both. Performance in the two word-form conditions was similar, with participants in the unmarked condition requiring an average of 2.4 blocks (range = 1–11) whereas participants in the marked condition required an average of 1.1 blocks (range = 1–2). The difference in means appears to be accounted for by two outliers in the unmarked condition: one who exceeded criterion in Block 4, and one who exceeded criterion in Block 11. An ANOVA conducted on the two word-form conditions indicated no significant difference.

In summary, a majority of the participants in this experiment were able to learn and when they learned they did so relatively quickly (within a block or so). There was no effect of word-form on the ease of learning. That is, participants learned equally well whether they were exposed to the marked or unmarked usage of the novel dimensional adjectives.

Accuracy in Comparative Learning

To examine the possibility of a semantic congruity effect in accuracy, we analyzed the percentage of correct responses in the first 5 blocks in a 2 (word-form) × 6 (stimulus pair) × 2 (term) mixed-design ANOVA. The data from only the first 5 blocks were used instead of the entire data set in order to maximize the possibility of finding an effect. Our reasoning was that, because the effect is often not found in highly accurate, overlearned perceptual comparisons (Petrusic & Baranski, 1989b), limiting our analysis to the initial learning stages would optimize our chances of finding an effect. However, performance even in these first 5 trials was very high (mean = 93% correct) and the analysis indicated no main effects or interactions. The stimulus pair by term interaction, which would signify a semantic congruity effect, did not approach significance, $F(5, 90) = 1.19$, $p = .32$. In addition, similar to the ease of acquisition results discussed earlier, there was no effect of word form, $F(1, 18) = .035$, $p = .85$, indicating that participants' learning was similar regardless of whether or not the novel dimensional words were marked explicitly as adjectives. A semantic congruity effect was not found in accuracy even when the analysis was limited to the very first trial block. Mean performance in Trial Block 1 was also very high, at 91% correct overall.

Categorization Test

Each participant's categorization test performance was classified into the three categories used in Experiment 1 (full categories, endpoints, and other) according to their overall performance on all trials at or above the comparative learning criterion discussed earlier. Similar to Experiment 1, a stimulus item was considered categorized if it was classified into one category 75% of the time or more and classified into the other category less than 25% of the time. According to these criteria, 17 participants were identified as showing some form of categorization performance (9 demonstrating full categorization and 8 demonstrating endpoint categorization), whereas 3 were classified as noncategorizers (2 from the marked condition, 1 from the unmarked condition). Because the learning criterion was met so early in the procedure by most participants, the experiment was divided into thirds (Trial Blocks 1–5, 6–10, and 11–15), and this partitioning was used as an independent variable in order to examine the possibility that the pattern of categorization performance changed during the course of the experiment. However, only data from the first 5 trial blocks (Trial Blocks 1–5) and the last 5 trial blocks (Trial Blocks 11–15) will be discussed. Analyses conducted on the middle blocks (Trial Blocks 6–10) indicated that the pattern of performance in these blocks did not differ from performance in the last 5 trial blocks. Thus, for the sake of brevity, the data from these trials will not be presented or discussed.

Data for the categorizers and noncategorizers for Trial Blocks 1–5 and 11–15 can be seen in Figures 8 and 9. As shown in these figures, participants identified as categorizers showed a strong crossover pattern indicative of categorization in both the initial third of the experiment and in the final third. However, although those identified as noncategorizers did not show such a pattern in the first third, they approached such a pattern in the final third (although with a great deal of variability). This conclusion was supported by further analyses of the data. Specifically, when the percentage of yes responses was submitted to a 2 (categorization status: categorizer vs. noncategorizer) × 2 (word-form) × 2 (Trial Blocks: 1–5 vs. 11–15) × 7 (stimulus position) × 2 (term) mixed-design ANOVA, it indicated, among other effects, three-way interactions among categorization status, stimulus position, and term, $F(6, 108) = 4.14$, $p < .01$, and among trial blocks, stimulus position, and term, $F(6, 108) = 2.87$, $p < .05$. The four-way interaction, however, did not reach significance, $F(6, 108) = 1.06$, $p = .39$. When the data from categorizers and noncategorizers in Trial Blocks 1–5 were analyzed separately, the stimulus position by term interaction was found to be significant only for the categorizers, $F(6, 96) = 54.22$, $p < .01$, and not for the noncategorizers, $F(6, 12) = .70$, $p = .66$. In contrast, when the data from the categorizers and noncategorizers in Trial Blocks 11–15 were analyzed separately, the stimulus position by term interaction was found to be significant for both the categorizers, $F(9, 96) = 68.34$, $p < .01$, and the noncategorizers, $F(6, 12) = 4.14$, $p < .05$. Thus, although the noncategorizers failed to meet our criteria for categorization, their classification of each object in the series was in the right direction by the end of the experiment.

In summary, the results of the categorization test indicate that when participants learned to accurately compare objects drawn from a series, they were likely to categorize these objects into two distinct groups even though they were given no categorization feedback at any time.

Reaction Time to Make Comparative Judgments

Because categorization might be crucial to the formation of a semantic congruity effect, the 17 categorizers' and 3 noncategorizers' reaction time data were analyzed separately (see Figure 10). The categorizers showed a semantic congruity effect in reaction time on the first 5 trial blocks but not in their performance on the last 5 trial blocks, as indicated by a significant three-way interaction among trial blocks, stimulus pair, and term, $F(5, 80) = 3.80$, $p < .01$. Thus, the semantic congruity effect changed over the course of the experiment. To examine the possibility that the effect in Trial Blocks 1–5 was limited to only the endpoints (because of their special treatment in training; see Experiment 1, Results), the endpoint pairs were removed from the analysis of these trial blocks, leaving only the 4 internal pairs. This analysis indicated that the semantic congruity effect was still significant, $F(3, 48) = 6.95$, $p < .01$. Individual analyses

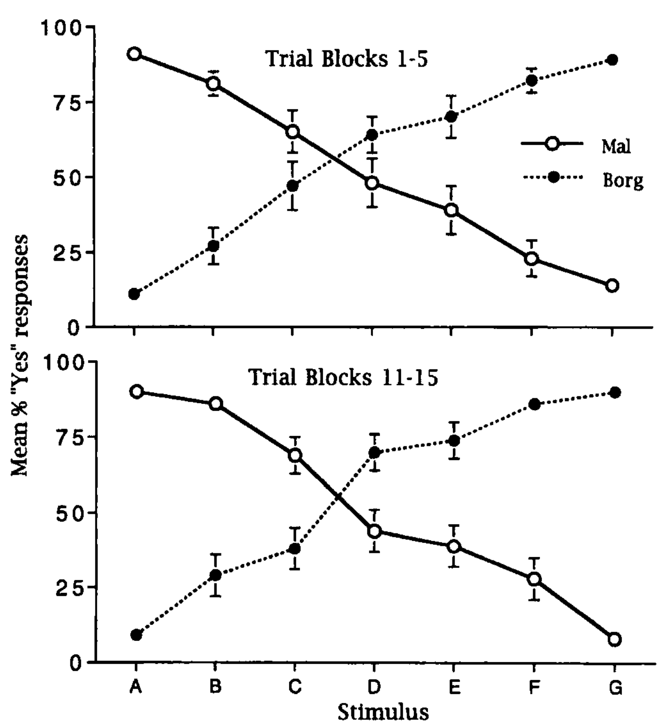

FIGURE 8. Experiment 2 categorization test results for categorizers, by trial blocks, stimulus, and term. *n* = 17. Vertical lines depict *SE*.

on each participant revealed that 14 of the 17 participants identified as categorizers displayed a significant semantic congruity effect in the first third of the experiment. By the last third of the experiment, the effect was significant in only 1 participant.

Unlike what was observed in the categorizers' data, there was no suggestion of a semantic congruity effect at either point in time in the noncategorizers' data. Analyses revealed no main effects or interactions. Similarly, when the noncategorizers were analyzed individually, none showed any indication of a semantic congruity effect in their reaction time performance at any point in time.

Summary

Similar to the results from the equivalent condition in Experiment 1, there was a strong relationship in this experiment among learning, formating of cate-

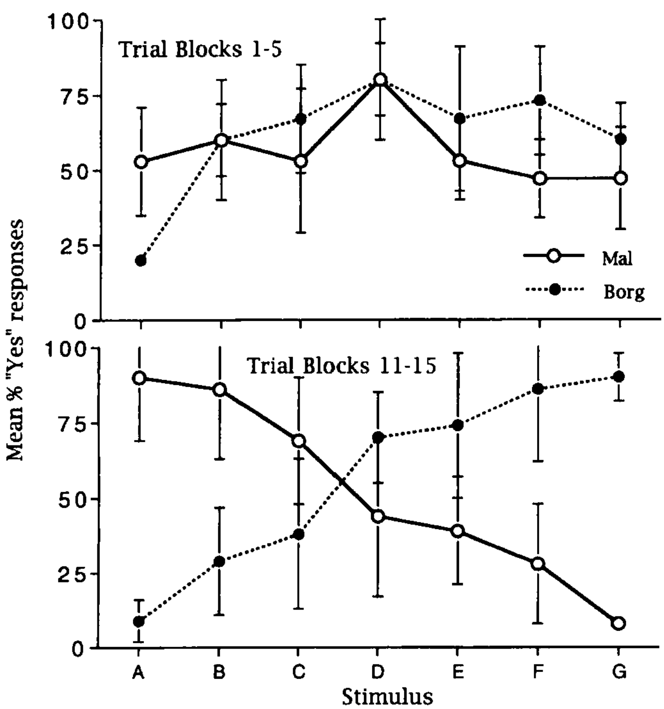

FIGURE 9. Experiment 2 categorization test results for noncategorizers, by trial blocks, stimulus, and term. $n = 3$. Vertical lines depict *SE*.

gories, and exhibiting of the semantic congruity effect in reaction time. Learners identified as showing some form of categorization also showed a significant semantic congruity effect in reaction time in the early stages of the experiment. Participants who did not meet our criteria for categorization (but who did show a tendency to categorize toward the end of the experiment) did not show a semantic congruity effect in reaction time at any time in the experiment. Although the semantic congruity effect in reaction time was robust among the categorizers, there was no indication of a semantic congruity effect in accuracy at any point in the experiment. The results were the same for both participants trained with the explicitly marked comparative terms *maller* and *borger* and those trained with the unmarked contrastive terms *mal* and *borg*.

Unlike the equivalent condition in Experiment 1, participants in the present stimulus condition were highly accurate in their comparative judgments from very early in training, performing correctly on 91% of the pairs in the first trial block. Two changes in the procedure may account for this difference. First, the

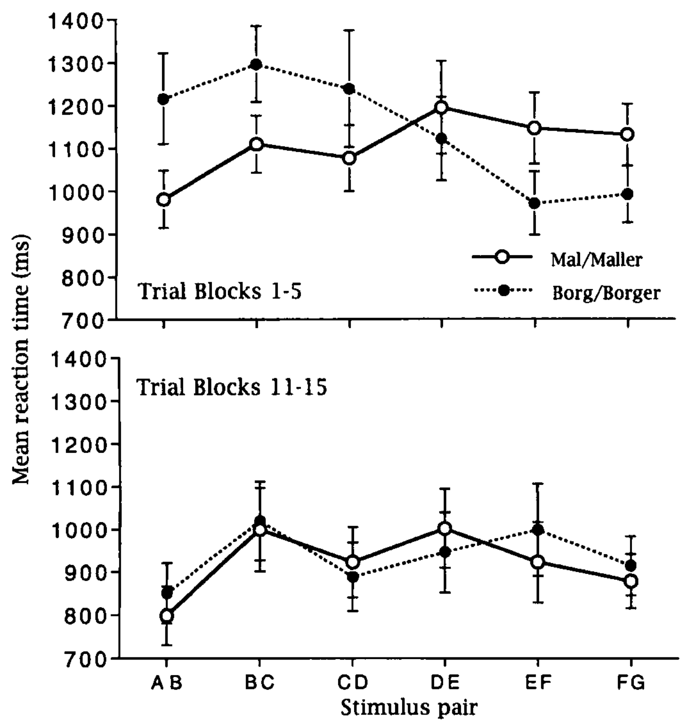

FIGURE 10. Experiment 2 comparative judgment test results for categorizers, by trial blocks, stimulus pair, and term. *n* = 17. Vertical lines depict *SE*.

internal features of the stimuli were removed, making the stimuli less complex. Second, both members of the pair to be compared were shown on the screen simultaneously (pair members were shown consecutively in Experiment 1). These changes may have made the difference in height more salient, thus allowing participants to quickly narrow their focus to a relevant dimension. If this is the case, then it may be that participants were simply translating *mal/ler* and *borg/er* to tall/er and short/er. This conclusion is compatible with the finding that the semantic congruity effect was present only in the initial trials. That is, previous research has indicated that the semantic congruity effect is sometimes not observed in the reaction time to make overlearned perceptual comparisons and is much more robustly observed in situations involving conceptual or symbolic comparisons (Petrusic & Baranski, 1989b).

In summary, the findings in Experiment 2 replicated the major results of Experiment 1 except for the noted differences in ease of acquisition and the finding that the semantic congruity effect was present only in the initial stages of the experiment. In Experiment 3, we replicated the other condition from Experiment 1 in which the number of learners was low and in which participants had access to only one of the two manipulated sources of information, the arbitrarily ordered stimuli/ordered-labeling condition. The same changes in procedure made for Experiment 2 (compared with Experiment 1) were also made in Experiment 3.

EXPERIMENT 3

Method

Participants

A total of 68 undergraduates ranging in age from 19 to 27 years were recruited from a midwestern U.S. university and were given extra credit in a psychology course for participation. The learning status of the participants was monitored, and data collection continued until there was a total of 20 learners—10 learners in the unmarked condition and 10 learners in the marked condition. There were roughly equal numbers of male and female learners in each condition. The data from the 48 nonlearners were not analyzed in detail and will be discussed only in terms of the relative ease of acquisition.

Materials and Apparatus

The arbitrarily ordered series used in this experiment was the same series of inanimate geometric shapes used in Experiment 1 (see Figure 1). Stimulus presentation and data collection were controlled by the application SuperLab (Cedrus, Inc.) running on a Power Macintosh 6100/60.

Procedure

The procedure in this experiment was identical to the procedure in Experiment 2 with two exceptions. First, as described earlier, the arbitrarily ordered stimulus series was used in this experiment. Second, during the training phase, stimuli were labeled using ordered labeling as described in Experiment 1 (see the Appendix for an example of one block).

Results and Discussion

Ease of Acquisition

As in Experiment 2, two measures of ease of acquisition will be discussed—the number of participants required in each condition to reach 10 learners and, for the learners, the block in which the learning criterion was met and maintained.

With regard to the number of participants required to reach 10 learners, in this experiment slightly fewer than 1 out of 3 (29%) participants reached the learning criterion. The 20 participants identified as learners averaged 88% correct on the last 5 trial blocks (range = 78–99%). The 48 participants identified as nonlearners averaged 54% correct on the last 5 trial blocks (range = 45–63% correct). Chance performance in this task was 50%. By this measure, and similar to the results of Experiment 1, the perceptual information available in the ordered stimulus series in Experiment 2 was more supportive of learning than the information available in the relative frequency of labeling in the present experiment. In the unmarked condition, 27% of the participants reached learning criterion, whereas 32% of the participants in the marked condition were able to reach criterion.

Examination of the second ease of acquisition measure indicated that the learners needed an average of 7.25 blocks to reach criterion, much longer than was necessary in Experiment 2. Performance in the two word-form conditions was somewhat different in this experiment, with participants in the unmarked condition requiring an average of 5.8 blocks (range = 2–13) and participants in the marked condition requiring an average of 8.7 blocks (range = 4–14). This difference, however, was not reliable, $F(1, 18) = 3.30, p = .09$.

Accuracy in Comparative Learning

There was no indication of a semantic congruity effect in accuracy. An analysis of the percentage of correct choices made by each participant in the first 5 blocks after criterion indicated only a main effect of stimulus pair, $F(5, 90) = 2.833, p < .05$. There was no main effect of word-form, $F(1, 18) = 1.01, p = .33$, and no interaction between stimulus position and term, $F(5, 90) = 1.34, p = .51$. The main effect of stimulus position resulted because participants were more accurate when asked about the end pairs than when asked about other pairs.

Categorization Test

Six participants in the marked condition were classified as showing some form of categorization (2 full categorization and 4 endpoint). Eight participants in the unmarked condition were classified as showing some form of categorization (5 full categorization and 3 endpoint). The remaining 6 participants were classified as noncategorizers. Because learning occurred relatively late in the procedure in this experiment (compared with performance in Experiment 2) and because participants varied widely in the number of blocks at or above criterion (range = 2–14 trial blocks; mean = 8.7), an average was calculated for each participant for all blocks at criterion. Thus, change within the course of the experiment will not be examined in this study as it was in Experiment 2.

The participants identified as showing some form of categorization (endpoint or full categorization) showed a strong crossover pattern, whereas participants identified as noncategorizers did not (see Figure 11). When these results were submitted to a 2 (categorization status) × 2 (word-form) × 7 (stimulus position) × 2 (term) mixed-design ANOVA, it revealed a marginal three-way interaction among categorization status, stimulus position, and term, $F(6, 96) = 1.91, p = .09$, as well as a main effect of stimulus position, $F(6, 96) = 5.82, p < .01$, and a significant interaction between stimulus position and term, $F(6, 96) = 19.47, p < .01$. Simple-effects analyses on the two categorization status subgroups indicated that the stimulus position by term interaction was highly significant for the categorizers, $F(6, 78) = 26.63, p < .01$, but did not reach significance within the noncategorizers group, $F(6, 30) = 2.03, p = .09$.

These results indicate that, similar to the findings in Experiment 2, although noncategorizers did not meet our criteria for categorization, they demonstrated some tendency toward organized performance. Specifically, 5 of the 6 noncategorizers met or almost met our criteria for categorization of one endpoint but not the other.

Reaction Time in Comparative Learning

The categorizers showed a semantic congruity effect in reaction time, whereas the noncategorizers did not (Figure 12). An ANOVA for a 2 (categorization status) × 2 (word-form) × 6 (stimulus pair) × 2 (term) mixed-design revealed a marginally significant three-way interaction among categorization status, stimulus pair, and term, $F(5, 80) = 2.211, p = .06$, and associated simpler effects (none of which involved word-form). When the categorizers were analyzed separately, the results revealed only a significant stimulus pair by term interaction, $F(5, 65) = 8.26, p < .01$. When the endpoint pairs were removed from the analysis, this semantic congruity effect remained significant, $F(3, 39) = 4.97, p < .01$, indicating the effect was a true semantic congruity effect and not an artifact produced by the differential treatment of the endpoints. In contrast, when the noncategorizers' average response times for all correct responses were analyzed separately, the analysis revealed no significant main effects or interactions, including a nonsignificant semantic congruity effect, $F(5, 25) = .56, p = .73$. Individual analyses on each participant revealed that 12 of the 14 participants identified as categorizers displayed a significant semantic congruity effect. Importantly, 2 of the 6 participants classified as noncategorizers also showed reliable semantic congruity effects in their reaction time to make comparative judgments. These 2 participants (along with 3 of the 4 other noncategorizers) showed some tendency to partition the series into two groups.

Summary

Once again, a relationship among learning, categorizing, and exhibiting the semantic congruity effect was observed. That is, even when the only information

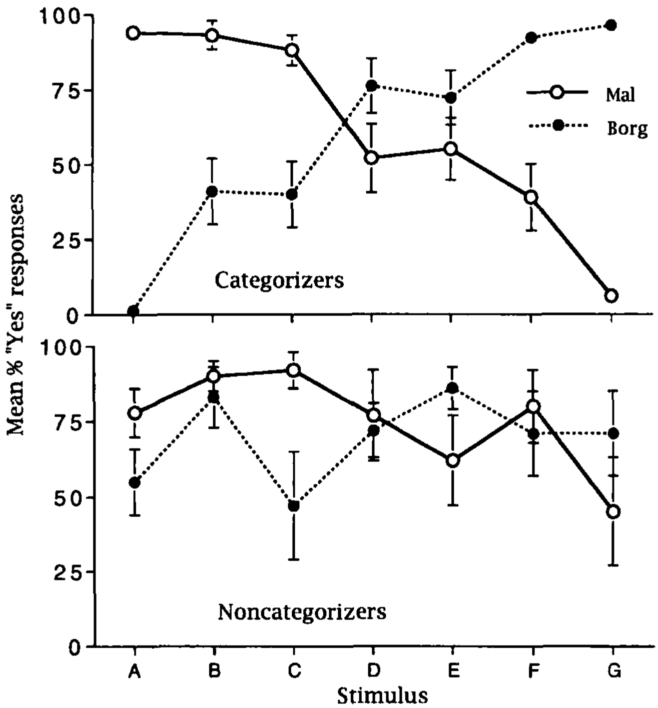

FIGURE 11. Experiment 3 categorization test results, by categorization status, stimulus, and term. n = 14 (top panel) and n = 6 (bottom panel). Vertical lines depict *SE*.

available to the participants concerning the structure of the series and the meaning of the words was in the relative frequency of labeling in the pair by pair comparison process, participants who learned the serial order were also likely to partition the series into categories and show a semantic congruity effect in comparison.

The "dimension" that participants learned in Experiment 3 was unlike any natural language dimension. That is, there was no perceptual information to order the stimulus series and no conceptual information to guide comparison above and beyond the information provided in the pair by pair comparison. Thus, it appears that the semantic congruity effect can emerge from the process of comparison alone with no underlying semantic information required (at least not as it is traditionally characterized). Although the literature indicates that the nature of the

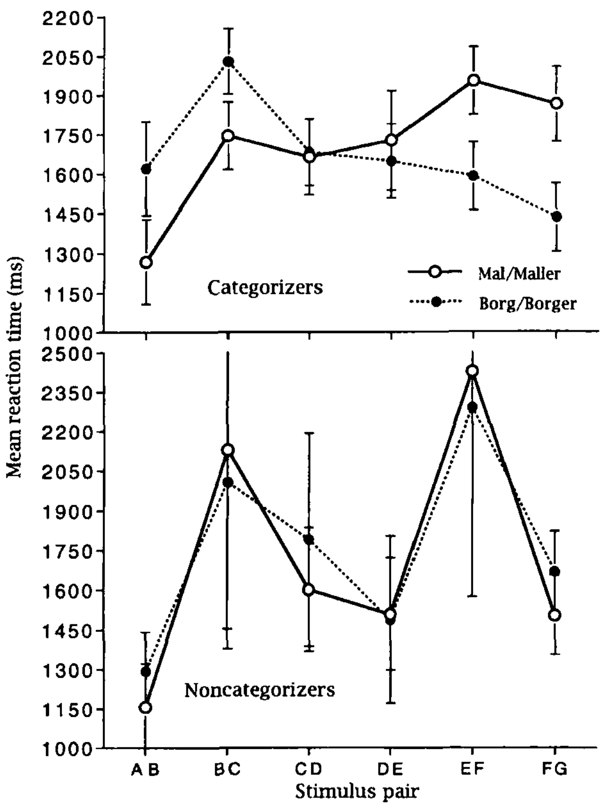

FIGURE 12. Experiment 3 comparative judgment test results, by categorization status, stimulus pair, and term: Reaction time data. n = 14 (top panel) and n = 6 (bottom panel). Vertical lines depict *SE*.

perceptual comparison and the participants' conceptual knowledge base can influence the semantic congruity effect, these influences are clearly not necessary to the formation of the effect.

We tested this hypothesis further in Experiment 4. Our reasoning was that if all that matters is knowing the relational meaning of the words—not the course of development, nor the particular characteristics of the stimuli or the training procedure—then even participants in the most informationally impoverished condition of Experiment 1, the arbitrarily ordered stimuli/equal-labeling condition, should show a semantic congruity effect in reaction time if they can be made to learn the meaning of the words. As discussed earlier, only 1 of the 10 participants in the analogous condition in Experiment 1 was classified as a learner. Therefore, in Experiment 4, we tested a small number of participants in the arbitrarily ordered stimuli/equal-labeling condition longitudinally in order to maximize the chances that at least some of the participants would learn, expecting that their pattern of acquisition would be similar to that of participants in the three prior experiments.

EXPERIMENT 4

Method

Participants

Three undergraduate women and 2 undergraduate men ranging in age from 19 to 22 were recruited from a midwestern U.S. university to serve in this experiment. All participants received monetary compensation for participation. Participation was terminated at the discretion of the experimenter or at the request of the participant (when the semester ended). Individual participation ranged from 9 to 15 sessions.

Materials, Apparatus, and Procedure

The materials and apparatus were the same as those used in the arbitrarily ordered stimuli/equal-labeling condition in Experiment 1. The procedure for this experiment was the same as that of Experiment 1 except that it was conducted longitudinally. Specifically, participants completed the three-session regimen of Experiment 1 repeatedly. After 9 sessions (12 training trial blocks and 3 speeded comparative-judgment tests) the comparative learning test and categorization test were discontinued for 3 participants who had reached asymptote in the comparative learning test. These people continued to participate, completing 1 and 2 additional speeded comparative-judgment tests, respectively, each preceded by 1 training block. The other 2 participants completed 15 sessions (20 training trial blocks and 5 speeded comparative-judgment tests), at which point the experiment was terminated because the semester ended. Sessions were conducted on consecutive days whenever possible, with no more than a 2-day interruption between sessions at any point in training.

Results and Discussion

Comparative Learning Test

In this experiment "learners" were defined as participants reaching the 70% correct criterion and maintaining that level of performance for at least 3 trial blocks. Three of the 5 participants met this criterion at some point in the procedure (2 women and 1 man). Figure 13A shows the overall percentage correct for the learners and nonlearners across trial blocks. The learners diverged from the nonlearners on Trial Block 9, when their performance jumped from 56% correct to 85% correct overall. The 2 nonlearners completed 20 training trial blocks each without reaching the learning criterion on a single trial block. The failure of the 2 nonlearners did not appear to be due to a lack of effort: Both participants appeared motivated and frequently voiced their frustration to the experimenter. Analysis of individual participants' data indicated that there were no semantic congruity effects in the accuracy of the participants' judgments for either the learners or nonlearners.

Categorization Test

As in the previous experiments, individual stimuli were considered to be categorized if an item was classified into one category 75% of the time or more and classified into the other category less than 25% of the time. Results of this test showed that participants who learned to make comparative judgments also categorized the stimuli. By the end of training, 2 learners strongly partitioned the series into two discrete categories. The 3rd learner never achieved full categorization, as defined earlier, but approached it with five out of seven items in the series placed in the appropriate categories. In order to capture the change in categorization behavior that occurred with training, we classified participants' categorization data for each trial block on a 5-point scale as follows: 1 = *categorization of one endpoint*; 2 = *categorization of both endpoints*; 3 = *categorization of both endpoints and stimulus B or F*; 4 = *categorization of both endpoints and stimulus B and F*; 5 = *full categorization.*

The learners diverged from nonlearners in their categorizations of individual objects on Trial Block 6, earlier than they diverged on the comparative learning test, reaching asymptote on Trial Block 10 (see Figure 13B). The nonlearners' categorization performance was very poor across all trial blocks, with no indication of any category organization by either nonlearner at any time during training.

Speeded Comparative-Judgment Test

As in Experiment 1, participants were tested on the speeded comparative judgment test after every 4th trial block. Figure 13C shows the number of learners and nonlearners exhibiting a significant semantic congruity effect in their reaction time. One learner produced a significant semantic congruity effect on the second speeded comparative judgment test after 8 training trial blocks. The other two learners showed significant semantic congruity effects on the third speeded comparative-judgment test after 11 training trial blocks. Neither nonlearner produced a semantic congruity effect in reaction time at any point in training.

Two of the learners completed 1 and 2 additional speeded comparative-

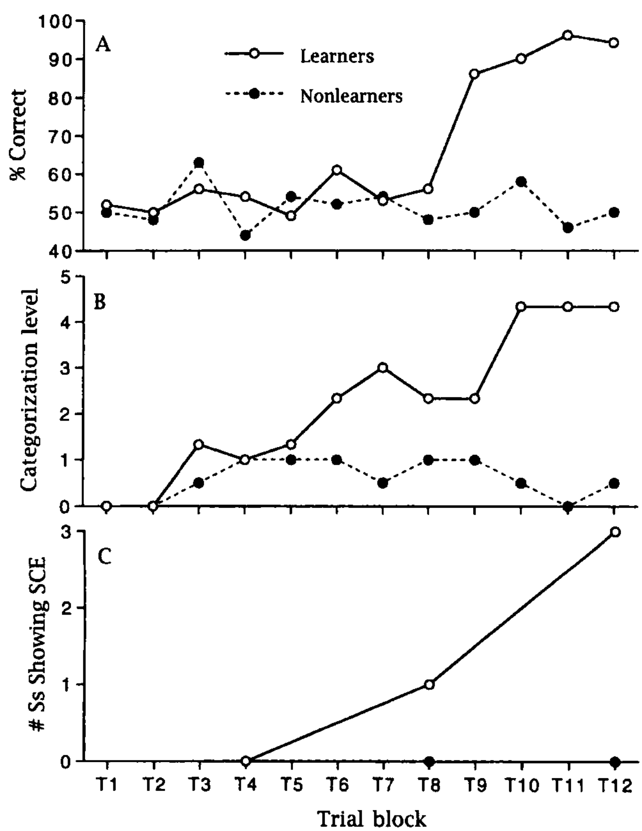

FIGURE 13. Results for learners and nonlearners in Experiment 4 are as follows: (A) comparative learning test, (B) categorization test, and (C) speeded comparative-judgment test.

judgment tests, respectively, each preceded by 1 block of training. On all of these additional trial blocks the participants continued to show significant semantic congruity effects.

Summary

In this experiment there was no perceptual information available in the stimulus series itself to indicate the relative position of items in the series, and there was no information about the ordering available from the frequency with which the individual items were explicitly labeled with each term. The only information available was in the pairwise comparison of individual objects, that is, the presentation of two objects with one labeled as *mal* or *borg*. This information over many blocks of training was sufficient for 3 of the 5 participants to learn the meaning of *mal* and *borg*, for categories to emerge, and for the development of a semantic congruity effect in speeded judgment. Thus, the results of this study indicate that at least some people can learn in this highly impoverished condition and that when they learn their performance resembles that of participants in other, less impoverished, conditions. However, because of the small number of participants, conclusions about the relative ease of learning (compared with the conditions tested in Experiments 2 and 3) cannot be drawn reliably, as it is unclear what percentage of adults would be capable of learning under such conditions, given a larger sample.

GENERAL DISCUSSION

The results of the present research can be summarized with regard to three related topics: (a) the adults' acquisition of novel relational words, (b) the relationship between learning relational words and categorization, and (c) the emergence of the semantic congruity effect. Each of these topics will be considered in turn. Finally, the questions that initially motivated this research, the relationship between adult and child word acquisition, are discussed.

Adult Acquisition of Novel Relational Words

At a global level, this work is about the acquisition by adults of novel relational words that refer to an ordered series. These experiments demonstrate that the type of information available to the learner greatly affects ease of learning. Not surprisingly, perceptual relations helped participants learn the series. More interestingly, the relative frequency with which individual items were labeled by contrasting terms also helped learners discover the series.

Across the experiments there were many nonlearners. This is not surprising. The real-world task presents learners with multiple sources of information through which to discover the meaning of dimensional terms. Even with these multiple sources, young word learners find these words difficult and take years to master them. In the present experiments, we sought to understand the structure of the learning task by subtracting these sources of information. Under these conditions, many did not learn. Given our goals for the present work, the key finding is that when people did learn the meanings of the two words that refer to opposing directions of difference on a linear order, they showed a semantic congruity effect in reaction time. Because of the exploratory nature of this work, the criterion for learning was loosely defined and determined in large part by a natural division in performance: Participants tended to perform below or near chance or well above chance. Future research should attempt to provide a more principled definition of "learning" and explore the relationship between learning and the other measures of interest more systematically. In addition, in the present work the characteristics that determined a participant's success or failure were not systematically examined and, thus, remain a question for future study.

Categorization and Novel Relational Word Meanings

The present results suggest that, regardless of the nature of the stimuli, the nature of the labeling, and whether participants were trained with marked or unmarked adjectives, if people learned to compare objects correctly, they were very likely to form categories. Specifically, out of 61 learners across four experiments, 51 (or 84%) partitioned the series into two categories. This formation of categories is interesting because categories were never taught to the participants but emerged as a consequence of acquiring relational word meaning. However, our results suggest that, although the formation of categories and the semantic congruity effect appear closely related, the formation of categories may not be necessary for the formation of a semantic congruity effect. Similarly, although many categorizers also mastered the relational word meanings, not all did. Categorization, therefore, is not sufficient or equivalent to the mastery of relational word meanings. These relationships between categorization and the semantic congruity effect are crucial because some explanations, notably Banks's discrete code model (1977) and its extension, Shoben's differential coding model (Cech, Shoben, & Love, 1990; Shoben, Cech, Schwanenflugel, & Sailor, 1989), view categorization as a critical component in comparative judgments and the semantic congruity effect. However, in these models categorical knowledge is often treated as interfering with relational judgments. That is, categorization is viewed as an obligatory process that must be gotten around if one is to make accurate comparative judgments. Our finding that the formation of categories occurs either just before or concurrently with learning dimensional adjectives raises the question of the causal relation between categories and relative meanings. We taught relative meanings and the participants formed categories. It seems possible that categories do not so much cause the semantic congruity effect as the processes that cause the semantic congruity effect also give rise to the emergence of categories.

One limitation of this study concerns classification of participants as categorizers or noncategorizers. In establishing criteria we set relatively high standards. Thus, some noncategorizers might have formed sufficient categories to yield a semantic congruity effect. Further research and formal modeling of the present results will be crucial to resolving these issues.

Origins and Explanations of the Semantic Congruity Effect

The results strongly indicate that the semantic congruity effect in reaction time does not depend critically on how dimensional meanings are taught but only on whether they are learned. In the present experiments, if participants learned the meaning of *mal* and *borg*, they produced a semantic congruity effect in reaction time. By far, the most interesting results in the present set of experiments are the semantic congruity effects found in conditions involving the arbitrarily ordered shape series because this continuum is very different from any ordered series tested before. Adults have, in the past, been trained to make comparative judgments about novel stimulus orders (see, for example, Moyer & Bayer, 1976; Polich & Potts, 1977; Woocher, Glass, & Holyoak, 1978). However, in those studies the participants were typically pretrained with previously familiar dimensional terms (i.e., Tom is taller than Joe; Joe is taller than Bill, etc.) or were pretrained to associate nonsense words with perceptual objects such as shapes varying in size. In Experiment 3, participants were not pretrained to associate the shapes with any known dimension. The only information available to learners was in the pairwise comparison of objects. There was neither quantitative information nor conceptual knowledge (of any traditional sort) underlying comparison of the shapes. What was being established in training, by the crudest and most impoverished means possible, and what learners must have possessed upon reaching criterion, was a mental representation of a linear order—a series formed by the pair by pair comparison of the stimulus items.

We believe that the explanation most compatible with the present set of findings is the evidence-accrual-based reference point model (Petrusic, 1992; see also Holyoak, 1978; Holyoak & Mah, 1982; Jamieson & Petrusic, 1975; Petrusic & Baranski, 1989a, 1989b). Reference point theory can account for comparative judgment findings with perceptual, symbolic, and remembered stimuli. In addition, according to Petrusic (1992), the reference point theory explicitly predicts the discriminability effects and speed–accuracy trade-off effects that have been reported in adult comparative judgments, whereas other theories, most notably the semantic coding theory, do not.

What we believe was happening in the experiments in this study, especially in the conditions and experiments involving the arbitrarily ordered shapes, was that participants were establishing reference points as they were constructing their representation of the linear order. That is, participants were taught that given A, B, and the probe *borg/er* say "B"; given B, C, and the probe *borg/er* say "C";

given C, D, and the probe *borg/er* say "D," and so on. Conversely, participants were taught that given A, B, and the probe *mal/ler* say "A"; given B, C, and the probe *mal/ler* say "B"; given C, D, and the probe *mal/ler* say "C," and so on. In this way, each stimulus was associated to some degree with each term and, in the context of that term, to the adjacent object on one side. Therefore, each object in the context of one term may have suggested the adjacent object that was in the direction of the named term. For example, "B" in the context of *mal* suggests "A." Thus, the pairwise comparison process was enough to establish two chains of associations that push attention in opposite directions toward context-sensitive reference points. What the present experiments make clear is that these reference points and the resulting semantic congruity effect can emerge in the process of acquisition with no pre-existing conceptual knowledge about the dimension in question required.

Children and Adults: Similarities and Differences

We began this research with the hypothesis that the semantic congruity effect in the speed of adult comparative judgments is the developmental product of how dimensional terms are learned. In the present set of experiments we varied the conditions of learning to be more or less like the conditions present in children's learning and examined adults' acquisition of novel dimensional adjectives. In general, the results suggest that the pattern of adults' learning of new dimensional terms is similar but not identical to the pattern of children's learning. The acquisition of relational meaning, the grouping of objects into categories, and, ultimately, the presence of a semantic congruity effect in reaction time were evident in the present artificial learning task. However, the pattern of acquisition displayed by adults was the same regardless of the specific information available to them in learning. In addition, a semantic congruity effect in accuracy was not reliably found in adult acquisition and thus is still robust only in children's acquisition.

The overall lack of semantic congruity effects in accuracy is consistent with previous research with adults (see Petrusic, 1992). Although the semantic congruity effect is quite robust in adult reaction time, it is quite elusive in accuracy and has been documented only once in a single experiment involving highly confusable stimuli in which speed and accuracy were equally important (Petrusic, 1992, Experiment 3). Petrusic (1992) has argued that the semantic congruity effect in accuracy is most likely to occur when responses are highly accurate (and response times are long). However, it is unclear how Petrusic's findings apply in the present case, as he was examining the psychophysical relationship between speed and accuracy in comparisons involving known words (nearer and farther) and very precisely controlled, highly confusable stimuli. In the experiments in this study, the relationship between speed and accuracy was not investigated because participants were in the process of acquiring the novel words and were presumably functioning at the highest level of performance possible. In addition, the stimuli in these experiments were very different from those used in Petrusic's studies. Although it is possible that the speed-accuracy trade-offs described by Petrusic (1992) may explain the lack of semantic congruity effects in accuracy in the present experiments, more research is needed before this conclusion can be accepted. What is clear, however, is that the semantic congruity effect in accuracy does not necessarily occur in adults' acquisition of novel dimensional adjectives.

In summary, the results suggest that, contrary to our initial hypotheses, the semantic congruity effect in adult reaction time for making comparative judgments does not depend on the particular aspects of the learning situation. Neither a perceptually ordered stimulus continuum nor differential labeling mattered except with regard to the ease of learning. In the experiments presented here, those participants who learned the meaning of our artificial dimensional terms, as impoverished as this meaning was, formed categories and also produced a semantic congruity effect. These results indicate that the semantic congruity effect is a process effect that can emerge in the act of comparison and is not dependent on conceptual knowledge or semantic information represented in long-term memory. Theories and models that attempt to explain the semantic congruity effect and related phenomena must be able to account for the present findings and, eventually, the developmental data that initially motivated this research. However, many questions remain about the relationship between the effect in children and the effect in adults. For example, little work exists that examines children's reaction time for making comparative judgments. Although it seems parsimonious to assume that the effect in children's accuracy is the same effect that is observed in adult reaction time, similarity in performance does not necessarily entail similarity in underlying mechanism. Thus, more work is required before the relationship between the two sets of findings will be completely understood.

REFERENCES

Audley, R. J., & Wallis, C. P. (1964). Response instructions and the speed of relative judgments. I. Some experiments on brightness discrimination. *British Journal of Psychology, 55*, 59–73.

Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11, pp. 101–159). San Diego: Academic Press.

Banks, W. P., Clark, H. H., & Lucy, P. (1975). The locus of the semantic congruity effect in comparative judgments. *Journal of Experimental Psychology: Human Perception and Performance, 104*, 35–47.

Banks, W. P., & Flora, J. (1977). Semantic and perceptual processes in symbolic comparisons. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 278–290.

Banks, W. P., White, H., Sturgill, W., & Mermelstein, R. (1983). Semantic congruity and expectancy in symbolic judgments. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 560–582.

Bierswisch, M. (1967). Some semantic universals of German adjectivals. *Foundations of Language, 3*, 1–36.

Cech, C. G. (1995). Is congruity due to encoding? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1275–1288.
Cech, C. G., & Shoben, E. J. (1985). Context effects in symbolic magnitude comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 299–315.
Cech, C. G., Shoben, E. J., & Love, M. (1990). Multiple congruity effects in judgments of magnitude. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1142–1152.
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.
Donaldson, M., & Wales, R. J. (1970). On the acquisition of some relational terms. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 235–278). New York: Wiley.
Ebeling, K. S., & Gelman, S. (1988). Coordination of size standards by young children. *Child Development, 59*, 888–896.
Ehri, L. (1976). Comprehension and production of adjectives and seriation. *Journal of Child Language, 3*, 369–384.
Ellis, S. H. (1972). Interaction of encoding and retrieval in relative age judgments: An extension of the "crossover" effect. *Journal of Experimental Psychology, 94*, 291–294.
Friend, K. E. (1973). Perceptual encoding in comparative judgments of race. *Memory and Cognition, 1*, 80–84.
Holyoak, K. J. (1978). Comparative judgments with numerical reference points. *Cognitive Psychology, 10*, 203–243.
Holyoak, K. J., & Mah, W. A. (1982). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology, 14*, 328–352.
Holyoak, K. J., & Patterson, K. K. (1981). A positional discriminability model of linear-order judgements. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1283–1302.
Jamieson, D. G., & Petrusic, W. (1975). Relational judgments with remembered stimuli. *Perception and Psychophysics, 18*, 373–378.
Klahr, D., & Siegler, R. (1978). The representation of children's knowledge. In H. Reese & L. P. Lipsitt (Eds.), *Advances in child development* (Vol. 12, pp. 61–116). San Diego: Academic Press.
Maratsos, M. (1973). Decrease in preschool children's understanding of the word big. *Child Development, 4*, 747–752.
Marks, D. F. (1972). Relative judgment: A phenomenon and a theory. *Perception & Psychophysics, 11*, 156–160.
Moyer, R. S., & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology, 8*, 228–246.
Petrusic, W. M. (1992). Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 962–986.
Petrusic, W. M., & Baranski, J. V. (1989a). Context, context shifts, and semantic congruity effects in comparative judgements. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures, mechanisms, and models. Proceedings of the XXIV International Congress of Psychology* (pp. 231–251). Dordrecht, The Netherlands: North-Holland.
Petrusic, W. M., & Baranski, J. V. (1989b). Semantic congruity effects in perceptual comparison. *Perception & Psychophysics, 45*, 439–452.
Polich, J. M., & Potts, G. R. (1977). Retrieval strategies for linearly ordered information. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 10–17.
Potts, G. R. (1972). Information processing strategies used in the encoding of linear order-

ings. *Journal of Verbal Learning and Verbal Behavior, 11*, 727–740.
Ryalls, B. O. (2000). Dimensional adjectives: Factors affecting children's ability to compare objects using novel words. *Journal of Experimental Child Psychology, 75*, 26–49.
Ryalls, B. O., Hall, M. J., Johnson, H. D., Reimer, J. F., & Anthis, K. S. (1999). *Maternal usage of dimensional adjectives in natural and elicited contexts.* Poster presented at the 1999 SRCD Biennial Meeting, Albuquerque, NM.
Ryalls, B. O., Winslow, E., & Smith, L. B. (1998). A semantic congruity effect in children's acquisition of high and low. *Journal of Memory and Language, 39*, 543–557.
Sera, M., & Smith, L. B. (1987). Categorical and relative interpretations of "big" and "little." *Cognitive Development, 2*, 89–112.
Shipley, W. C., Coffin, J. I., & Hadsell, K. C. (1945). Reaction time in judgments of color preference. *Journal of Experimental Psychology, 35*, 206–215.
Shoben, E. J., Cech, C. G., Schwanenflugel, P. J., & Sailor, K. M. (1989). Serial position effects in comparative judgments. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 273–286.
Shoben, E. J., & Wilson, T. L. (1998). Categorization in judgments of relative magnitude. *Journal of Memory and Language, 38*, 94–111.
Sinclair-de-Zwart, H. (1969). Developmental psycholinguistics. In D. Elkind & J. H. Flavell (Eds.), *Studies in cognitive development* (pp. 315–366). New York: Oxford University Press.
Smith, L. B., Cooney, N., & McCord, C. (1986). What is "High"? The development of reference points for "High" and "Low." *Child Development, 57*, 583–602.
Smith, L. B., Ratterman, M. J., & Sera, M. (1988). "Higher" and "lower": Comparative and categorical interpretations by children. *Cognitive Development, 3*, 341–357.
Trabasso, T. (1977). The role of memory as a system of making transitive inferences. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 333–366). Hillsdale, NJ: Erlbaum.
Wallis, C. P., & Audley, R. J. (1964). Response instructions and the speed of relative judgments. II. Pitch discrimination. *British Journal of Psychology, 55*, 133–142.
Woocher, F. D., Glass, A. L., & Holyoak, K. J. (1978). Positional discriminability in linear orderings. *Memory & Cognition, 6*, 165–173.

APPENDIX

Examples of an Ordered-Labeling Training Trial Block and an Equal-Labeling Training Trial Block

Ordered-labeling block		Equal-labeling block	
C F Mal	D A Borg	F E Mal	C A Mal
B D Mal	C D Mal	D G Borg	D A Mal
E B Borg	G B Borg	B A Mal	E B Mal
D A Borg	F C Borg	D G Borg	F G Borg
E F Mal	C G Mal	F A Mal	E A Mal
D B Borg	G D Borg	E G Borg	B G Borg
D E Mal	F E Borg	C G Borg	E F Borg
E D Borg	C E Mal	F A Mal	F G Borg
G E Borg	D E Mal	C A Mal	B C Borg

(Appendix continues)

APPENDIX (continued)

Ordered-labeling block		Equal-labeling block	
B A Borg	A B Mal	F G Borg	E D Mal
F D Borg	B F Mal	C B Mal	B F Borg
C B Borg	D G Mal	D F Borg	B A Mal
E G Mal	A C Mal	E A Mal	E G Borg
A G Mal	D C Borg	F G Borg	C E Borg
E G Mal	F A Borg	B A Mal	B D Borg
E A Borg	C G Mal	B E Borg	C D Borg
F G Mal	B D Mal	D C Mal	D G Borg
B G Mal	A F Mal	B A Mal	F G Borg
A E Mal	B G Mal	F D Mal	E C Mal
B C Mal	E C Borg	C G Borg	C B Mal
F D Borg	C A Borg	D A Mal	E F Borg
B E Mal	D F Mal	E G Borg	D B Mal
G A Borg	E A Borg	C F Borg	C A Mal
F C Borg	G F Borg	B A Mal	D E Borg
C A Borg	F B Borg	D C Mal	F C Mal

Note. The stimulus pair presented is indicated by capital letters, and the question asked ("Which is *mal*?" or "Which is *borg*?") is indicated by the word *mal* or *borg*. In this Appendix the correct choice is the stimulus item on the left side. In the experiments the side of the correct choice (left or right) was counterbalanced.

Manuscript received March 23, 1999
Revision accepted for publication July 29, 1999