Predicting Taxonomic and Thematic Relational Responding

J. Grayson Osborne and John Heath
Utah State University

Pairs of pictures were classified by the authors and others as related by identity (A–A), basic taxonomy (A–B), superordinate taxonomy (A–C), or by theme (A–D). Two-choice matching-to-sample trial types were composed of these same picture pairs in which the sample was common to the two stimulus pairs in each configuration and, together with the sample, each comparison exemplified one of the relations in the picture pair; that is: A(AB), A(AC), A(AD), A(BC), A(BD), and A(CD). In five experiments, for each picture pair, college students classified the relation (as taxonomic or thematic) and rated its strength (Exps 1, 3) or its similarity (Exp 4); others matched to sample the foregoing trial types only (Exps 2, 5), or they classified and rated, too (Exp 3). With exceptions, students classed most pairs as the authors did. They also collectively ordered relational strengths from (1) identity, (2) basic taxonomy, and (3) theme, to (4) superordinate taxonomy based, in part, on the similarity of sample and comparison. Subjects chose the comparisons of the more strongly related picture pairs in the matching-to-sample task on 90 percent or more of the configurations. Subjects' selections in two-choice, matching-to-sample configurations using natural stimuli may be based on existing stimulus control topographies such as those exhibited by ratings of the relations in a configuration.

Humans are said to relate natural stimuli characteristically at different ages (Denney, 1974; Denney & Ziobrowsky, 1972; Smiley & Brown, 1979). Early in life it is supposed that they relate natural stimuli thematically, while later during school and adult years they relate them taxonomically. In still later life they may again more often relate natural stimuli thematically (Smiley & Brown, 1979). The evidence for these changes in relational responding between and among natural stimuli consists of aggregated responses averaged across subjects from two general, but distinct, procedures, one in which a subject is given picture (or real object) stimuli and asked to sort them. If arrangements result in which cause appears to relate the stimuli (e.g., horses and riders together), the subject is said to relate the stimuli themati-

cally. If the subject categorizes (e.g., horses are placed with other horses), the subject is said to relate the stimuli taxonomically.

A second procedure—more structured—is to engage the subject in the generic conditional discrimination known as matching to sample (cf. Sidman & Tailby, 1982) using similar stimuli. In this task the subject is presented a picture (or real object) sample and two picture (or real object) comparison stimuli, one that relates to the sample taxonomically and the other that relates to the sample thematically (Osborne & Calhoun, 1998). Selection of a comparison related to the sample taxonomically or a comparison related to the sample thematically permits the inference of a stimulus–stimulus (or just stimulus) relation between sample and comparison that is taxonomic or thematic, respectively (Osborne & Calhoun, 1998). Thus, matching-to-sample configurations such as these contain two potential stimulus relations, only one of which controls comparison selection on any trial. Using this task, the data less typically support the age-related hypothesis. Rather, selection of comparison stimuli relates to contextual variables such as reinforcement for one type of selection (Bauer & Mandler, 1989; Osborne & Koppel, 2001), observation of a model differentially selecting comparisons of one kind or another (Smiley &

Reprints can be obtained from J. Grayson Osborne, Department of Psychology, Utah State University, 2810 Old Main Hill, Logan, UT 84322-2810. John Heath is now in the PhD program in Psychology at Auburn University. The authors thank Leslie Engstrom (now at Southern Illinois University), Loetta Wood, Michael Warde, Ryan Ward (now at University of New Hampshire), Michael Evans, and Marcia Phillips for their help in constructing the stimuli and running Experiments 4 and 5.

Brown, 1979), receipt of instructions that suggest taxonomic selection (Osborne & Calhoun, 1998; Waxman & Namy, 1997), or control by specific contextual stimuli (Osborne & Koppel, 2001). Moreover, individual subjects exhibit differential frequencies of taxonomic or thematic selections (Dunham & Dunham, 1995; Osborne & Calhoun, 1998). Finally, individual matching-to-sample configurations may differentially evoke thematic or taxonomic comparison selections by most subjects (Osborne & Calhoun, 1998). It is this last finding that is of interest here.

Responding on taxonomic/thematic tasks has been of considerable theoretical significance to those involved in developmental and cognitive psychology (e.g., Markman & Hutchinson, 1984; Markman, 1991). The so-called thematic preference has been attributed to children's cognitive competence at early age—the implications being that they are either not taxonomically competent in early life (Bruner, Goodnow, & Austin, 1956; Inhelder & Piaget, 1964; Vygotsky, 1962) or that their performance fails to evidence such competence (Gelman, 1978; Gelman & Greeno, 1989). It is now understood that even very young infants are taxonomically competent and that taxonomic/thematic responding is context dependent (Bauer & Mandler, 1989; Waxman & Namy, 1997). But what is (are) the context(s) upon which such responding depends? As mentioned above, on one level, reinforcement contingencies, modeling, and instructions all provide contexts. On another level, contexts can also be provided by arbitrary stimuli that enable taxonomic or thematic selections on individual configurations (Osborne & Koppel, 2001). Absent all of these contexts there remains only the individual trial configuration and that too appears to provide its own context (Osborne & Calhoun, 1998). Unanswered to this point is what it is about a particular matching-to-sample stimulus configuration that may evoke taxonomic or thematic selection, absent other contextual variables.

Which stimulus relation in a particular configuration controls comparison selection may in part depend on how similar a comparison is to the sample (Fenson, Cameron, & Kennedy, 1988). Additionally, the two stimulus relations in a configuration may be differentially relevant or familiar in terms of the subjects' experiences. One way to capture these experiences is to have

subjects respond to the individual stimulus relations that comprise a matching-to-sample configuration. If subjects differentially respond to the individual stimulus relations in a task that is not matching to sample, such differential responding might presage systematic responding to the entire matching-to-sample configuration.

Operationalizing the foregoing, in a conditional discrimination that contains two stimulus relations—one that exemplifies taxons and one that exemplifies themes—it should be possible to predict selection based on what subjects say, via ratings, is the strongest of the two relations. Subject ratings of the strengths of the stimulus relations comprising the configuration of stimuli on any matching-to-sample trial may identify particular stimulus control topographies (McIlvane & Dube, 1992). If the stimulus control topographies of the stimulus relations in the two tasks cohere, then differential responding in one should predict differential responding in the other. If comparison selection in matching to sample were shown to relate to prior behavior another contextual variable, the stimulus relations comprising the individual configuration, would be shown to account for individual variability in this conditional discrimination.

Three Types of Taxonomic Matching to Sample

A second purpose of the present studies was to construct matching-to-sample trial types to reflect potentially different strengths of taxonomic and thematic relations. To that end we reflect on some of the different kinds of taxonomic relations that may be compared in a matching-to-sample format utilizing natural stimuli.

*Matching to sample in trial types containing A as sample and A as one comparison: A(AX).*¹ There are at least three types of taxonomic relations between natural stimuli: In the first, a stimulus is related to itself (cf. Sidman & Tailby, 1982). For example, a picture of a cow is related to the identical picture of the

¹ The first letter (A) denotes the sample in a matching-to-sample task. The first letter in the parentheses denotes a comparison in a matching-to-sample task. The second letter in the parentheses denotes a second comparison. X = a generic comparison.

cow. The conditionally discriminated response in which the subject matches a stimulus with itself, in its generalized form exemplifies the logical property called *reflexivity* (Sidman & Tailby, 1982). Given that the features of the two stimuli that are matched in these trial types are identical, responding on this type of matching to sample has been called *identity matching* (but cf., Sidman & Tailby, 1982). As trial types that exemplify A(AX) are presented early and often to children, in situations where more than one controlling relation exists (see below), the stimuli comprising the A–A relation should control subjects' selections over other types of relation (A–X). Furthermore, the stimulus relation operationalized by an A sample and an A comparison should be judged by subjects as most strongly related. The high strength of A–A relations may result from the perceptual similarity (indeed, the perceptually identical nature) of the two stimuli comprising the relation (Baldwin, 1992; Fenson et al., 1988).

What the functionally controlling stimulus relation is in the matching-to-sample trial types exemplified by A(AX) may be difficult to discern because there are several stimulus control topographies that exist in these trial types. One source of stimulus control relates the identity of the stimuli. That is, though the stimuli also exemplify a taxon (e.g., cow1 to cow1), the taxonomic relation need not be functional. Alternately, it is also theoretically possible that the stimulus control topography in this trial type is taxonomic exclusive of the identical features of the two stimuli (cf. Lowenkron, 1998). A third possibility is that the stimulus control topography in this trial type is if ... then (Carter & Werner, 1978), that is, if A1, select B1. In this last instance, the subject's comparison selections could be under the control of neither the identical physical features, nor the taxonomic relation exemplified by the stimuli, but only the if ... then relation established by the contingencies in a conditional discrimination. Thus, even given the subject's consistent selection of the A–A relation, the coexisting stimulus control topographies make the reason for the selection on the A(AX) trial type ambiguous.

Matching to sample in trial types containing A as sample and B as a basic taxonomic comparison: A(BX). A second type of taxonomic matching to sample occurs when the stimuli that are related taxonomically are not

identical, but are of the same basic kind. For example, a picture of a guernsey cow is related to a picture of a different cow, say a holstein. The cows have the same general shape but are different in the details (e.g., horns, color, aspect, etc). In these trial types also, multiple stimulus control topographies make the functionally controlling stimulus relation ambiguous because the stimuli can be related in the same several ways: perceptual similarity, basic taxonomy, and if ... then. Given that perceptual similarity between the sample (A) and one comparison stimulus (B) in configurations exemplifying A(BX) matching to sample is still high, A–B stimulus relations should also be judged strong, but, perhaps not as strong as relations based on perceptual identity (A–A).

Matching to sample in trial types containing A as sample and C as a superordinate taxonomic comparison: A(CX). A third type of taxonomic matching to sample results when the sample (A) in a stimulus configuration relates to one comparison stimulus at a superordinate level (C); for example, the subject matches a cow (sample) with a horse (comparison) since both are mammals. In A(CX) trial types, perceptual similarity is lower than in A(AX) and A(BX) trial types. The controlling relations in A(CX) trial types should be comparatively less strong than those either in A(AX) or in A(BX) because there is less perceptual similarity between sample (A) and comparison (C) (e.g., cow as sample and whale as comparison).

In A(CX) configurations as well there are multiple stimulus control topographies, and responding could occur based on if ... then, similarity, and taxonomic relations between the stimuli. Similarity, while lower than in the other kinds of taxonomic matching, may not be absent. For example, a subject may be more likely to relate two stimuli that exemplify animals if both have eyes or any minimally comparable body parts, especially if the remaining stimuli in the trial type have no features in common with the sample (e.g., they are plants) (Baldwin, 1992).

Thematic, A(DX), Matching to Sample

A fourth kind of matching to sample involves a relation between sample (A) and comparison (D) based on theme, exemplified by sample (dog) and comparison (bone), or sample (horse) and comparison (saddle). As noted above, this type of relation between stimuli has been called

thematic by developmental psychologists. In A–D relations there is no perceptual similarity between the sample and the comparison. Superficially, it would seem that the strength of these relations may be somewhat lower than those of taxonomic relations where similarity may relate the stimuli. However, humans are exposed early and often to stories that relate such stimuli.

There is little information from a behavior analytic framework about the stimulus control topographies extant in the foregoing trial types. It is possible that a hierarchy of stimulus control topographies is exemplified by the foregoing trial types in which stimulus relations exemplifying identity, basic taxonomy, superordinate taxonomy, and theme are contrasted.

Given the foregoing rationales, in the experiments that follow, it is determined: 1) whether subjects consistently order the frequencies of their taxonomic classifications, e.g., from A–A to A–D; 2) whether subjects consistently order the strength ratings of the relations, e.g., from A–A to A–D; and 3) whether prior or subsequent subject ratings of relation strengths in picture pairs exemplifying A–A, A–B, A–C, and A–D relations relate to prior or subsequent relation selections in matching-to-sample trial types composed of the stimulus pairs with common samples, that is, A(AB), A(AC), A(AD), A(BC), A(BD), and A(CD).

EXPERIMENT 1

The purpose of this experiment was to test subject responses to 80 picture pairs in which the pairs exemplified an identity relation (A–A), a basic taxonomic relation (A–B), a superordinate taxonomic relation (A–C), or a thematic relation (A–D). Subjects were to classify each pair of pictures as exemplifying either a taxonomic or a thematic relation and then to rate the strength of the relation between the pictures on a 5-point Likert scale. It was predicted that taxonomic classifications would be most frequent during A–A pair types and less frequent in each pair type thereafter, i.e., A–B, A–C, and A–D. It was further predicted that relation strength ratings would be strongest for A–A pairs and less strong for each pair type thereafter, i.e., A–A, A–B, A–C, and A–D.

METHOD

Subjects

Subjects ($N = 27$) volunteered from a freshman psychology class at Utah State University. Mean age of the subjects was 20.7 yrs. Subjects were offered extra credit for their participation in the experiment. The experiment was conducted in a classroom of the university and utilized computer controlled audio visual equipment.

Stimuli

Stimuli consisted of colored pictures of common objects selected from "Corel Magic 200,000" software (© 1997 Corel Corporation and Corel Corporation, Limited). Each trial consisted of a slide showing a pair of pictures horizontally arrayed, one picture that would eventually function as a sample (in matching to sample), and a second that would eventually function as a comparison that could be matched to the sample. An example of two of the stimulus pairs and the matching-to-sample configuration they could become is shown in Figure 1.

Eighty slides were produced of these pairs of pictures. Each slide was composed of a to-be-sample stimulus and a second stimulus related to the first in one of four ways, in the authors' judgment, i.e., by identity with it (A–A); by a basic taxonomic relation with it (A–B); by a superordinate taxonomic relation with it (A–C); or by a thematic relation with it (A–D). Table 1 lists all of the picture stimuli.

Procedure

When all the subjects were in the classroom each was given an informed consent form to read and sign, and each completed personal information on optically readable forms (Scantron F262).

Pre-training. After the experimenter (JH) introduced himself, subjects were told they were to decide whether a slide illustrated a taxonomic or a thematic relation after which they were to rate the strength of the relation between the stimulus pairs in each slide by marking their Scantron forms. The subjects were then shown a slide of two dogs, an example of a taxonomic relation. The experimenter stated, "It can be

Figure 1. A representation of the two kinds of tasks in the experiments. In the upper panels the two stimulus pairs (cow and milk bottles; cow and horse) represent two of the arrangements, observed in single pairs, when the subjects judged the kind of relation between the stimuli and then rated the strength of the relation. Cow and milk bottles depict a thematic relation (A1-D1); cow and horse depict a superordinate taxonomic relation (A1-C1) (see Table 1). The bottom panel depicts a matching-to-sample configuration composed of the common stimulus of the two stimulus relations in the top panels as sample, and the remaining stimuli (milk bottles and horse) as comparisons. Each matching-to-sample configuration in the studies was composed in this fashion from combining two of the stimulus pairs with common samples (Table 1).

classed taxonomic because the two dogs belong to the same class; that is, they are both dogs." Next, a slide of a dog and cat was shown. The experimenter stated, "Here is a sample of another taxonomic relation. Again they are in the same class, i.e., they are both mammals." The next slide showed a dog and a piece of meat. The experimenter stated, "Here is an ex-

ample of a thematic relation. It is a thematic relation because there is a theme connecting the two pictures; that is, dogs eat meat." Next, the subjects were shown three further example slides—lock and key, cigarette and cigarette, and cigarette and pipe. After showing each slide the subjects were asked to raise their hands if they thought the relation was taxonomic or thematic. This procedure was used to indicate whether or not the subjects understood the difference between taxonomic and thematic selection. All subjects indicated in each case that they understood.

The penultimate introductory slide was a representation of the Scantron form so that precise recording instructions could be given, while the ultimate introductory slide was an example of the instructions that would be on each slide to indicate to the subjects on which line of the Scantron to record their selection. Each slide showed a stimulus from Table 1, column A (1A–20A) to the left of the slide, with a stimulus from the same or one of the other columns horizontally aligned with the stimulus from column A to the right side of the slide (Figure 1, upper panels). The picture pair used approximately the top half the slide. Below were instructions on how to mark the Scantron form, e.g. "Use Scantron line #1: 1 = Taxonomic; 2 = Thematic. Scantron line #2: 1 = Not related; 2 = Very poorly related; 3 = Poorly related; 4 = Well related; 5 = Very well related. Following this, slides of the pairs of picture stimuli were presented. Each was presented for approximately 20 seconds or until it was apparent that all the subjects had filled in their Scantron forms. The order of the picture pairs was randomized with restrictions that the picture on the left of the pair could not appear twice consecutively and the exemplified relation could occur no more than twice, consecutively. The picture pairs were placed into "Corel Presentation 8" in the randomized order. Including 9 for pre-training, there were a total of 89 slides. At the end of the presentation, the Scantron forms were collected by an assistant together with the consent forms. The subjects were thanked for their participation and dismissed.

Results

Classification. The classification results are shown in Table 2. There was 100% judgment

Table 1. The 80 Stimuli Used in Experiments 1–3.

Sample No.	Sample Stimulus (A)	Basic Taxon Taxon (B)	Superordinate Taxon (C)	Theme (D)
1	Cow 1	Cow 2	Horse	Milk bottles
2	Frog 1	Frog 2	Turtle	Fly
3	Ladybug 1	Ladybug 2	Beetle	Lettuce head
4	Chicken 1	Chicken 2	Duck	Nest with egg
5	Apple 1	Apple 2	Pear	Fruit tree
6	Fish 1	Fish 2	Shrimp	Fisherman with pole
7	Soda can 1	Soda can 2	Soda bottle	Hamburger
8	Bread loaf 1	Bread loaf 2	Doughnut	Jar of jam
9	Green pepper 1	Green pepper 2	Onions	Shish kabob on plate
10	Car 1	Car 2	Pickup truck 1	Engine part
11	Sailboat 1	Sailboat 2	Row boat	Life jacket
12	Motorcycle 1	Motorcycle 2	Moped	Motorcycle helmet
13	Snowmobile 1	Snowmobile 2	Sled	House window c snow
14	Semi-truck 1	Semi-truck 2	Pickup truck 2	Gas can
15	Light bulb 1	Light bulb 2	Candle c holder	Book and spectacles
16	Violin 1	Violin 2	Harp	Orchestra conductor
17	Toaster 1	Toaster 2	Kitchen stove	Homemade jam
18	Upholstered chair 1	Upholstered chair 2	Sofa	Television set
19	Backpack 1	Backpack 2	Coolers	Tent
20	Hair comb 1	Hair comb 2	Hair brush	Mirror

in favor of the taxonomic class given A–A pairs, 99.4% judgment in favor of the taxonomic class given the A–B pairs, 92.0% judgment in favor of the taxonomic class given the A–C pairs, and 88.3% judgment in favor of the thematic

class given A–D pairs. There was one stimulus pair (19A–19C) from the taxonomic pairs (i.e., A–A, A–B, A–C) in which a slight majority of subjects judged the relation as thematic (19A = backpack; 19C = two coolers). Also, A–D

Table 2. Classification Frequencies for the 80 Stimulus Pairs (Experiment 1).

Pair No.	A		B		C		D	
	Tax	Theme	Tax	Theme	Tax	Theme	Tax	Theme
1	27	0	27	0	25	2	0	27
2	27	0	26	1	26	1	2	25
3	27	0	27	0	25	2	0	27
4	27	0	27	0	26	1	3	24
5	27	0	27	0	25	2	5	22
6	27	0	27	0	25	2	0	27
7	27	0	26	1	27	0	15	12
8	27	0	27	0	25	2	5	22
9	27	0	27	0	23	4	10	17
10	27	0	27	0	25*	1	1	26
11	27	0	27	0	27	0	1	26
12	27	0	27	0	26	1	1	26
13	27	0	27	0	24	3	3	24
14	27	0	27	0	27	0	0	27
15	27	0	26	1	26	1	0	27
16	27	0	27	0	25	2	0	27
17	27	0	27	0	24	3	1	26
18	27	0	27	0	25	2	3	24
19	27	0	27	0	13	14	6	21
20	27	0	27	0	27	0	2	25
Sums	540	0	537	3	496	43	63	477
%	100	0	99.4	0.6	92.0	8.0	11.7	88.3

* One response to 10C was omitted by a subject.

pair, 7A–7D (7A = soda bottle; 7D = hamburger), was judged taxonomic by a majority of subjects.

Relation strength. Table 3 shows the mean relation strength for each of the stimulus pairs in the four pair types as judged by the subjects. A–A pairs were rated 5 by all but one subject (Mean of Means = 4.99; $S^D = 0.17$); A–B pairs slightly lower (Mean of Means = 4.63; $S^D = 0.51$); and A–C pairs slightly lower yet (Mean of Means = 3.64; $S^D = 0.82$). Thematic (A–D) pairs were rated about the same as the superordinate pairs (Mean of Means = 3.68; $S^D = 1.13$) but there was greater variation in the A–D ratings.

A one-way, within-subjects, analysis of variance of relation strength was run with each type of stimulus pair as factors. There was a significant effect of type of stimulus pair on the mean relation strength ratings, Wilks' $\lambda = .20$; $F(3, 532) = 688.98$; $p < .001$; $\eta^2 = .79$. Follow-up, pair-wise comparisons showed that mean rela-

tion strength ratings differed between all pairs of means ($p < .001$) except for A–C (superordinate) and A–D (thematic).

Discussion

The stimulus pairs were more likely to be judged as taxonomic the more similar the stimuli in the pair appeared to be. Subjects' judgments of relation strength decreased in an orderly fashion, as apparent similarity between the stimuli in the pairs decreased across the kinds of taxonomic relations with thematic strength ratings being no different than those of the superordinate stimulus pairs.

EXPERIMENT 2

Next, we were interested in how another group of subjects would respond to the foregoing picture pairs when those picture pairs comprised a conditional discrimination task,

Table 3. Mean Strength Ratings for Stimulus Pairs (Experiment 1).

Stimulus Pair (An-An)	Mean Strength Rating	Stimulus Pair (An-Bn)	Mean Strength Rating	Stimulus Pair (An-Cn)	Mean Strength Rating	Stimulus Pair (An-Dn)	Mean Strength Rating
A1-A1	5.00	A1-B1	4.48	A1-C1	3.44	A1-D1	4.59
A2-A2	5.00	A2-B2	4.59	A2-C2	3.30	A2-D2	3.93
A3-A3	5.00	A3-B3	4.59	A3-C3	3.92	A3-D3	2.67
A4-A4	5.00	A4-B4	4.54	A4-C4	3.33	A4-D4	4.63
A5-A5	5.00	A5-B5	4.81	A5-C5	3.81	A5-D5	4.67
A6-A6	5.00	A6-B6	4.55	A6-C6	3.26	A6-D6	4.44
A7-A7	5.00	A7-B7	4.22	A7-C7	4.44	A7-D7	3.44
A8-A8	5.00	A8-B8	4.74	A8-C8	3.11	A8-D8	4.07
A9-A9	5.00	A9-B9	4.96	A9-C9	3.33	A9-D9	3.44
A10-A10	5.00	A10-B10	4.55	A10-C10	3.65	A10-D10	3.44
A11-A11	5.00	A11-B11	4.78	A11-C11	4.07	A11-D11	3.59
A12-A12	4.85	A12-B12	4.55	A12-C12	3.70	A12-D12	4.33
A13-A13	5.00	A13-B13	4.81	A13-C13	3.37	A13-D13	2.41
A14-A14	5.00	A14-B14	4.67	A14-C14	3.67	A14-D14	3.74
A15-A15	5.00	A15-B15	4.59	A15-C15	3.41	A15-D15	3.00
A16-A16	5.00	A16-B16	4.37	A16-C16	3.96	A16-D16	3.89
A17-A17	5.00	A17-B17	4.78	A17-C17	3.41	A17-D17	2.44
A18-A18	5.00	A18-B18	4.48	A18-C18	4.07	A18-D18	3.33
A19-A19	5.00	A19-B19	4.74	A19-C19	3.33	A19-D19	3.78
A20-A20	5.00	A20-B20	4.74	A20-C20	4.15	A20-D20	3.74
Means	4.99	Means	4.64	Means	3.63	Means	3.67

i.e., matching to sample, of different trial types. Each matching-to-sample configuration was composed of three stimuli: one, a sample, and two others—comparisons. In some trial types one comparison related to the sample taxonomically and the other related to the sample thematically (e.g., Figure 1, lower panel), and in other trial types, with the common sample, each comparison exemplified a different type of taxonomic relation. The taxonomic relations were identity (A–A), ordinate (A–B), and superordinate (A–C). We could then observe whether there were systematic selections of identical, basic, or superordinate, taxonomically related comparisons and/or thematically related comparisons in the various trial types. Finally, it could be determined if the strength ratings generated in Experiment 1 related to comparison selection in this experiment.

METHOD

Subjects

Subjects ($N = 10$) were again volunteers from beginning undergraduate psychology classes

at Utah State University. Mean age of the subjects was 20.8 yrs. Subjects were again offered extra credit for their participation in the experiment.

Stimuli

Two hundred forty slides were composed from the stimuli described in Table 1. Each slide, a matching-to-sample configuration, contained three pictures, one of which—a sample stimulus—was always a stimulus from Column A of Table 1. The sample stimulus was placed at the top center of the slide. The two comparison stimuli, which were all possible combinations of pairs of stimuli from Columns A through D of Table 1 from the same row as the sample, were placed at the bottom left and right hand corners of the slide (Figure 1, lower panel). Together the three stimuli formed a rough equilateral triangle with the sample as apex. Trial types were A(AB), A(AC), A(AD), A(BC), A(BD), and A(CD), where A was the sample, A was an identical taxonomic comparison, B was a basic taxonomic comparison, C

Table 4. Frequencies of Selection of Comparisons (Experiment 2).

Sample (A)	CO:AB	CO:AC	CO:AD	CO:BC	CO:BD	CO:CD
A1	A1=18* B1=2	A1=17* C1=3	A1=12* D1=8	B1=19* C1=1	B1=14 D1=6	C1=4 D1=16*
A2	A2=19* B2=1	A2=18* C2=2	A2=16* D2=4	B2=15* C2=5	B2=12* D2=8	C2=8 D2=12*
A3	A3=20* B3=0	A3=18* C3=2	A3=16* D3=4	B3=18* C3=2	B3=16* D3=4	C3=16* D3=4
A4	A4=13* B4=7	A4=20* C4=0	A4=15* D4=5	B4=20* C4=0	B4=16 D4=4	C4=7 D4=13*
A5	A5=18* B5=2	A5=16* C5=4	A5=16* D5=4	B5=16* C5=4	B5=15* D5=5	C5=8 D5=12*
A6	A6=19* B6=1	A6=17* C6=3	A6=15* D6=5	B6=19* C6=1	B6=12* D6=8	C6=3 D6=17*
A7	A7=18* B7=2	A7=20* C7=0	A7=16* D7=4	B7=18 C7=2	B7=10 D7=10	C7=13* D7=7
A8	A8=18* B8=2	A8=17* C8=3	A8=16* D8=4	B8=17* C8=3	B8=13* D8=7	C8=5 D8=15*
A9	A9=18* B9=2	A9=16* C9=4	A9=16* D9=4	B9=16* C9=4	B9=16* D9=4	C9=12 D9=8
A10	A10=17* B10=3	A10=17* C10=3	A10=15* D10=5	B10=17* C10=3	B10=18* D10=2	C10=18* D10=2
A11	A11=14* B11=6	A11=17* C11=3	A11=19* D11=1	B11=18* C11=2	B11=18* D11=2	C11=16* D11=4
A12	A12=18* B12=2	A12=19* C12=1	A12=16* D12=4	B12=18* C12=2	B12=17* D12=3	C12=5 D12=15*
A13	A13=17* B13=3	A13=20* C13=0	A13=18* D13=2	B13=20* C13=0	B13=19* D13=1	C13=14* D13=6
A14	A14=20* B14=0	A14=18* C14=2	A14=18* D14=2	B14=18* C14=2	B14=17* D14=3	C14=15 D14=5
A15	A15=19* B15=1	A15=19* C15=1	A15=16* D15=4	B15=16* C15=4	B15=15* D15=5	C15=14* D15=6
A16	A16=17* B16=3	A16=18* C16=2	A16=18* D16=2	B16=20* C16=0	B16=17* D16=3	C16=6 D16=14
A17	A17=19* B17=1	A17=16* C17=4	A17=16* D17=4	B17=16* C17=4	B17=15* D17=5	C17=10 D17=10

(Continued, p. 64)

Table 4 (continued).

Sample (A)	CO:AB	CO:AC	CO:AD	CO:BC	CO:BD	CO:CD
A18	A18=19* B18=1	A18=14* C18=6	A18=14* D18=6	B18=15* C18=5	B18=16* D18=4	C18=17* D18=3
A19	A19=18* B19=2	A19=17* C19=3	A19=17* D19=3	B19=16* C19=4	B19=14* D19=6	C19=2 D19=18*
A20	A20=17* B20=3	A20=16* C20=4	A20=16* D20=4	B20=17* C20=3	B20=16* D20=4	C20=15* D20=5
Totals	A=356 (89.0%) B=44 (11.0%)	A=350 (87.5%) C=50 (12.5%)	A=321 (80.3%) D=79 (19.7%)	B=349 (87.2%) C=51 (12.8%)	B=306 (76.5%) D=94 (23.5%)	C=208 (52.0%) D=192 (48.0%)
% Hits	100%	100%	100%	95%	85%	80%

*Matching-to-sample outcome predicted by mean strength of relation rating from Table 3. The eight cells of the table that are outlined are not predicted by the results of Table 3.

was a superordinate taxonomic comparison, and D was a thematic comparison.

Procedure

Each slide was given a number visible to the subjects in the bottom left hand corner that corresponded with the order in which it was presented. Each configuration was presented twice, once with each comparison stimulus in the left or right position. The slides were sorted into a quasi-random order with the criteria that no two slides with the same sample followed one another and that not more than two instances of one trial type could occur successively. The slides were then inserted into Corel Presentation 8 to present to the subjects via computer, projector, and screen.

The experimenter (JH) told the subjects, "In a few minutes you are going to see a number of slides with pictures on them. Each slide contains three pictures, one on the top of the slide and two at the bottom which are side by side. Your task is to look first at the picture at the top, second at the two bottom pictures, then to decide which one of the bottom two pictures goes best with the top picture. In front of you, you have a Scantron form. If you think that the left picture goes best with the picture at the top, you will mark the Scantron in column 1. If you think the right picture goes best with the

picture at the top, you will mark the Scantron in column two. The number in the bottom left hand corner of the slide will tell you which line on the Scantron to mark your selection."

The subjects were then shown an example of how the slides would be presented—a picture of a dog as sample, a picture of a similar dog as the left side comparison and a picture of a piece of meat as the right side comparison. They were then told by the experimenter, "Here is a slide with three pictures on it. If you think the dog at the bottom goes best with the dog at the top, then you would mark column 1 on line 1 of your Scantron. You will note the number 1 on the slide to indicate the Scantron line. If you think that the meat goes best with the dog at the top, then you would mark column two on your Scantron."

The slides were then presented automatically on a large screen in the front of the classroom at intervals of 15 s. With the presentation of slides 60, 120, and 180 the subjects were reminded of what line of the Scantron they should be marking.

Results

Table 4 shows the collected responses of the subjects to each trial type (columns) for each sample (rows). The table shows that for trial types A(AB), A(AC), and A(AD) where the

sample stimulus could be matched with itself, that is, A was a sample and the same A was a comparison, subjects most frequently chose A. Whenever an arbitrary match was required—on trial types A(BC), A(BD), and A(CD)—subjects chose B when B and C were comparisons; chose B when B and D were comparisons; and chose C and D about equally often when C and D were comparisons (see column totals).

Table 4 also shows the results of relating relation selection with the ratings that were produced in Experiment 1. Each of the two stimulus pairs that comprised a configuration in the matching-to-sample trials in Experiment 2 had a relation strength rating from Experiment 1 (Table 3). For example, in configuration A1(A1B1) in Table 4, the choice of relation A1–A1 was predicted from the difference between the mean strength ratings in Table 3, where the strength rating for A1–A1 = 5.0 and for A1–B1 = 4.48. Each cell in Table 4 was compared with two cells in Table 3 in this way. The more strongly rated relation in Table 3 was used to predict relation selection in the matching-to-sample task. If the more strongly rated relation in Table 3 was selected more frequently in Table 4, then the cell in Table 4 contains an asterisk. Cells in Table 4 in which the less strongly rated relation in Table 3 was selected more frequently are outlined and boldfaced. All cells in Table 4 contain asterisks for trial types A(AB), A(AC), and A(AD); 19 of 20 cells for trial types A(BC); 17 of 20 cells for trial types A(BD); and 16 of 20 cells for trial types A(CD).

Discussion

Overall, the more strongly related of the two relations was selected on 112 of the 120 arrays (93.33%). Configurations in the A(CD) trial type were composed of one thematic (A–D) and one (superordinate) taxonomic relation (A–C). Given that similarity between sample and comparison is only an issue for A–C but not for A–D relations in this trial type, did strength ratings relate to comparison selection here as well? In fact, eight of 10 more strongly rated A–D relations were selected while the less strongly rated A–C relation was selected on two others. The more strongly rated A–C relation was selected eight times, while the less strongly rated A–D relation was selected twice. If relation strength only reflected similarity between

sample and comparison, there should have been no correct A–D selections. The foregoing notwithstanding, subjects appeared to match on the basis of perceptual similarity of sample and comparison on trial types where that was possible.

EXPERIMENT 3

The prior two experiments used different subjects for the strength ratings (Exp 1) and the matching to sample (Exp 2). In this experiment subjects experienced both procedures from the prior experiments. Half the subjects experienced the prior conditions in the order of Experiments 1 and 2 (classify and rate; match to sample) and half experienced those conditions in the reversed order (match to sample; classify and rate).

METHOD

Subjects

College students were selected as they were in the prior two experiments. Thirty-seven completed the two procedures in two sittings of about an hour each. Their average age was 18.6 yrs. Subjects who completed both procedures were paid \$10 because the time commitment was such that initial attempts to get subjects to sit solely for extra credit failed. Eighteen subjects matched to sample first and classified and rated strength second, while 19 subjects classified and rated first and matched to sample second.

Procedures

The procedures of this experiment were the same as those of the prior two experiments.

Results Classification.

The classification results for all subjects are shown in Table 5. There was 100% judgment in favor of the taxonomic class given A–A pairs, 96.3% judgment in favor of the taxonomic class given the A–B pairs, 75.1% judgment in favor of the taxonomic class given the A–C pairs, and 93.2% judgment in favor of the thematic class given A–D pairs. There was one stimulus

Table 5. Categorization Frequencies for the 80 Stimulus Pairs (Experiment 3).

Pair No.	A		B		C		D	
	Tax	Theme	Tax	Theme	Tax	Theme	Tax	Theme
1	37	0	37	0	31	6	1	36
2	37	0	35	2	28	9	1	36
3	37	0	36	1	33	4	1	36
4	37	0	35	2	29	8	6	31
5	37	0	35	2	32	5	6	31
6	37	0	36	1	25	12	1	36
7	37	0	37	0	33	4	7	30
8	37	0	36	1	26	11	3	34
9	37	0	36	1	25	12	11	26
10	37	0	36	1	33	4	1	36
11	37	0	36	1	35	2	1	36
12	37	0	35	2	30	7	2	35
13	37	0	36	1	24	13	2	35
14	37	0	35	2	35	2	1	36
15	37	0	34	3	25	12	0	37
16	37	0	34	3	26	11	0	36
17	37	0	37	0	20	17	1	36
18	37	0	36	1	33	4	2	35
19	37	0	35	2	4	33	3	34
20	37	0	36	1	29	8	0	37
Sums	740	0	713	27	556	184	50	690
%	100	0	96.3	3.7	75.1	24.9	6.8	93.2

pair (A19–C19) from the taxonomic pairs in which a majority of subjects classified the relation as thematic (19A = backpack; 19C = two coolers).

Relation strength. Table 6 shows the mean relation strength ratings for each of the stimulus pairs in the four categories for all subjects. All but 4 A–A pairs were rated 5 (Mean of Means = 4.99; $S^D = 0.03$); A–B pairs were rated lower (Mean of Means = 4.37; $S^D = 0.09$); and A–C pairs were rated lower yet (Mean of Means = 3.48; $S^D = 0.37$). Thematic (A–D) pairs were rated higher than the superordinate pairs (Mean of Means = 3.59; $S^D = 0.59$).

Two, one-way, within-subjects, analyses of variance of relation strength were conducted with each type of stimulus pair as factors—one for each group. For the group that matched first and classified and rated second, there was a significant effect of type of stimulus pair on the mean relation strength ratings, Wilks' $\lambda = .21$; $F(3, 355) = 453.2$; $p < .001$; $\eta^2 = .79$.

Follow-up, pair-wise comparisons showed that mean relation strength ratings differed between all pairs of means ($p < .001$). For the group that classified and rated first, and matched to sample second, there was also a significant effect of type of stimulus pair on the mean relation strength ratings, Wilks' $\lambda = .18$; $F(3, 376) = 559.52$; $\eta^2 = .82$. Follow up pair-wise comparisons showed that mean relation strength differed between all pairs of means ($p < .001$), except A–C and A–D ($p < .07$).

Table 7 shows the collected comparison responses of the subjects in both groups to each trial type (columns) for each sample (rows) in the matching-to-sample task. The table shows that for trial types A(AB), A(AC), and A(AD), where the sample stimulus could be matched with itself, subjects mostly chose A. Whenever an arbitrary match was required—A(BC), A(BD), and A(CD)—subjects more frequently chose B when B and C were comparisons; more frequently chose B when B and D were com-

Table 6. Mean Strength Ratings for Stimulus Pairs (Experiment 3).

Stimulus Pair (An-An)	Mean Strength Rating	Stimulus Pair (An-Bn)	Mean Strength Rating	Stimulus Pair (An-Cn)	Mean Strength Rating	Stimulus Pair (An-Dn)	Mean Strength Rating
A1-A1	5.00	A1-B1	4.43	A1-C1	3.21	A1-D1	4.37
A2-A2	5.00	A2-B2	4.38	A2-C2	2.86	A2-D2	3.84
A3-A3	5.00	A3-B3	4.40	A3-C3	3.50	A3-D3	2.63
A4-A4	5.00	A4-B4	4.27	A4-C4	2.86	A4-D4	4.37
A5-A5	5.00	A5-B5	4.54	A5-C5	3.56	A5-D5	4.47
A6-A6	5.00	A6-B6	4.32	A6-C6	3.46	A6-D6	4.47
A7-A7	5.00	A7-B7	4.25	A7-C7	4.24	A7-D7	3.00
A8-A8	5.00	A8-B8	4.51	A8-C8	3.08	A8-D8	3.79
A9-A9	5.00	A9-B9	4.50	A9-C9	3.54	A9-D9	3.58
A10-A10	5.00	A10-B10	4.32	A10-C10	3.72	A10-D10	3.37
A11-A11	5.00	A11-B11	4.48	A11-C11	3.81	A11-D11	3.79
A12-A12	5.00	A12-B12	4.40	A12-C12	3.26	A12-D12	4.05
A13-A13	4.97	A13-B13	4.26	A13-C13	3.27	A13-D13	2.68
A14-A14	5.00	A14-B14	4.37	A14-C14	3.34	A14-D14	3.42
A15-A15	5.00	A15-B15	4.32	A15-C15	3.24	A15-D15	2.79
A16-A16	4.91	A16-B16	4.35	A16-C16	3.87	A16-D16	3.89
A17-A17	4.97	A17-B17	4.40	A17-C17	3.19	A17-D17	2.81
A18-A18	5.00	A18-B18	4.18	A18-C18	3.97	A18-D18	3.26
A19-A19	4.89	A19-B19	4.30	A19-C19	3.41	A19-D19	3.74
A20-A20	5.00	A20-B20	4.27	A20-C20	4.11	A20-D20	3.63
Means	4.99	Means	4.37	Means	3.48	Means	3.59

parisons; and when C and D were comparisons chose C and D about equally (see column totals).

Table 7 also shows the results of predictions based on the mean strength of relation ratings that these same subjects produced. These predictions were generated as they were in Experiment 2 and are presented the same way. All cells in Table 7 contain asterisks for trial types A(AB), A(AC), and A(AD), and A(BC); for 17 of 20 cells for trial type A(BD); and for 17 of 20 cells for trial type A(CD).

Individual Ss

Figure 2 shows the numbers of selections of the most strongly rated stimulus relation for all 18 subjects who matched to sample first and rated second, for each of the six matching-to-sample trial types. The same analysis of relation strength that was applied collectively above was applied to the individual subjects. As each of the 20 configurations was presented twice within each trial type, there were 40 pos-

sible selections of the strongest relation per trial type, across the six trial types. The horizontal midline in each panel in Figure 2 identifies 20 selections of the most strongly rated relations which, in this two-choice task, would be chance (50%). The last panel to the right of the bottom row depicts means for each trial type for the group.

Viewing the panels as wholes, 11 of the 18 subjects' selections of the stronger relation were greater than 50% on all the trial types. Second, vertically dividing each panel in half permits a comparison of A(AX) and arbitrary matching-to-sample trial types—A(BC), A(BD), A(CD). Fourteen of the 18 subjects' selections of the stronger relation on the A(AX) trial types were greater than 50% and 13 of the 18 subjects' selections of the stronger relation on the arbitrary matching-to-sample trial types were also above 50%. Third, selection of the stronger relation on the A(CD) trial type was above 50% for 17 of the 18 subjects. Fourth, collectively, the panel depicting the means shows that, for each trial type, selection of the strongest rela-

Table 7. Frequencies of Comparison Selections (Experiment 3).

Sample (A)	CO:AB	CO:AC	CO:AD	CO:BC	CO:BD	CO:CD
A1	A1=64* B1=10	A1=68* C1=6	A1=58* D1=16	B1=69* C1=5	B1=49 D1=25	C1=13 D1=61*
A2	A2=68* B2=6	A2=73* C2=1	A2=66* D2=8	B2=70* C2=4	B2=56* D2=18	C2=23 D2=51*
A3	A3=72* B3=2	A3=73* C3=1	A3=67* D3=7	B3=67* C3=7	B3=67* D3=7	C3=59* D3=15
A4	A4=63* B4=11	A4=74* C4=0	A4=60* D4=14	B4=73* C4=1	B4=53 D4=21	C4=13 D4=61*
A5	A5=66* B5=8	A5=72* C5=2	A5=56* D5=18	B5=71* C5=3	B5=63* D5=11	C5=22 D5=52*
A6	A6=72* B6=2	A6=73* C6=1	A6=59* D6=15	B6=70* C6=4	B6=53 D6=21	C6=10 D6=64*
A7	A7=72* B7=2	A7=69* C7=5	A7=57* D7=17	B7=51* C7=23	B7=53* D7=21	C7=55* D7=19
A8	A8=72* B8=2	A8=70* C8=4	A8=61* D8=13	B8=73* C8=1	B8=46* D8=28	C8=17 D8=57*
A9	A9=71* B9=3	A9=67* C9=7	A9=68* D9=6	B9=64* C9=10	B9=62* D9=12	C9=40 D9=34
A10	A10=70* B10=4	A10=70* C10=4	A10=66* D10=8	B10=70* C10=4	B10=66* D10=8	C10=61* D10=13
A11	A11=65* B11=9	A11=74* C11=0	A11=60* D11=14	B11=70* C11=4	B11=59* D11=15	C11=47 D11=27
A12	A12=66* B12=8	A12=72* C12=2	A12=56* D12=18	B12=71* C12=3	B12=49* D12=25	C12=20 D12=54*
A13	A13=67* B13=7	A13=69* C13=5	A13=68* D13=6	B13=69* C13=5	B13=66* D13=8	C13=60* D13=14
A14	A14=69* B14=5	A14=73* C14=1	A14=65* D14=9	B14=73* C14=1	B14=62* D14=12	C14=52 D14=22
A15	A15=69* B15=5	A15=73* C15=1	A15=61* D15=13	B15=71* C15=3	B15=60* D15=14	C15=46* D15=28
A16	A16=64* B16=10	A16=63* C16=11	A16=63* D16=11	B16=62* C16=12	B16=53* D16=21	C16=33 D16=41*
A17	A17=72* B17=2	A17=63* C17=11	A17=65* D17=9	B17=63* C17=11	B17=55* D17=19	C17=41* D17=33

(Continued, p. 69)

Table 7 (continued). Frequencies of Comparison Selections.

Sample (A)	CO:AB	CO:AC	CO:AD	CO:BC	CO:BD	CO:CD
A18	A18=70* B18=4	A18=59* C18=15	A18=60* D18=14	B18=52* C18=22	B18=51* D18=23	C18=62* D18=12
A19	A19=70* B19=4	A19=67* C19=7	A19=62* D19=12	B19=67* C19=7	B19=48* D19=26	C19=4 D19=70*
A20	A20=66* B20=8	A20=59* C20=15	A20=56* D20=18	B20=60* C20=14	B20=59* D20=15	C20=63* D20=11
Totals	A= 1,368 (92.4%) B= 112 (7.6%)	A=1,380 (93.2%) C=100 (6.8%)	A=1,233 (89.9%) D=247 (10.1%)	B=1,336 (90.3%) C=144 (9.7%)	B=1,130 (76.3%) D=350 (23.7%)	C=741 (50.1%) D=739 (49.9%)
% Hits	100%	100%	100%	100%	92.5%	92.5%

*Matching-to-sample outcome predicted by mean strength rating from Table 6. The six cells of the table that are outlined are not predicted by the results of Table 6.

tion was above 50%. A binomial test was run to determine that overall predictions of subjects' comparison selections did not differ from chance (i.e., 120 of 240 predictions). The numbers of correct predictions statistically exceeded 120 for all subjects ($p < .01$).

Figure 3 shows the numbers of selections of the stronger relation for the 19 subjects who classified and rated first and matched to sample second, for each of the six trial types. Viewing the panels as wholes, 17 of the 19 subjects' selections of the more strongly rated relations were greater than 50% on all the trial types. All 19 subjects' selections of the more strongly rated relations on the A(AX) trial types were above 50%—although some are just barely above—and 17 of the 19 subjects' selections of the more strongly rated relations on the arbitrary matching-to-sample trial types were also above 50%. Third, selection of the more strongly rated relation on the A(CD) trial type was above 50% for 18 of the 19 subjects. Fourth, collectively, the panel depicting the means shows that, in each trial type, selection of the stronger relation was above 50%. A binomial test was run to determine that correct predictions of subjects' comparison selections did not differ from chance (i.e., 120 of 240 predictions). The numbers of correct predictions statistically exceeded 120 for all subjects ($p < .01$).

Discussion

Overall, the more strongly rated of the two relations was selected on 114 of the 120 arrays (95.0%).

In the A(CD) trial type, where the A–D relation was rated more strongly it was selected on nine of 12 occasions, while the less strongly related A–C relation was selected on the other three occasions. Where the A–C relation was more strongly rated, it was selected eight times on eight occasions, with no selections of the less strongly rated A–D relation. Again, if rated relation strength only reflected similarity between sample and comparison, there should have been no A–D selections.

The relation strength analysis worked somewhat less well when applied to individual subjects. This was because there were numerous occasions where relation strength ratings were judged identically for a subject for the two relations comprising a configuration. Under these circumstances a subject would earn the maximum only if her/his selection was controlled on one presentation of a particular configuration by one relation and on the next presentation of the same configuration was controlled by the other relation. Nevertheless, on most of the trial types for most of the subjects, selection was of the stronger relation on significantly more than 50% of the configurations.

Figure 2. Frequency distributions of the numbers of comparison selections that agreed with predictions based on the most strongly rated relation across the six matching-to-sample trial types for the 18 subjects in Experiment 3 who matched to sample first and classified and rated second. The panel marked Means is composed of the means of all 18 subjects for each trial type.

There was considerable similarity between the results of Experiment 3 and those of Experiment 1. With respect to classification

(Tables 2 and 5), in both experiments all subjects classified all A–A pairs as taxonomic. Subjects sporadically classified A–B pairs as

Figure 3. Frequency distributions of the numbers of comparison selections that agreed with predictions based on the most strongly rated relation across the six matching-to-sample trial types for the 19 subjects in Experiment 3 who classified and rated first and matched to sample second. The panel marked Means is composed of the means of all 19 subjects for each trial type.

thematic and did so more in Experiment 3 than in Experiment 1, where 17 versus three configurations showed at least one selection as the-

matic. But, collectively, all A-B pairs were classed as taxonomic. Almost all A-C pairs were occasionally classified as thematic, but

in both experiments 19 of 20 were classed as taxonomic by a majority of subjects. A19–C19 was an exception in both experiments. In Experiment 1, 19 of 20 and in Experiment 3, 20 of 20 A–D pairs were classed as thematic.

In both experiments, mean strength ratings of the stimulus pairs were ordinaly identical: A–A pairs strongest; A–B pairs next; A–D pairs third; and the A–C pairs last. In both Experiments 2 and 3, in matching-to-sample trial types A(AB), A(AC), and A(AD), comparison selection was controlled by A–A (Tables 4 and 7). There was overlap from Experiments 2 to 3 with respect to selection of the less strongly rated relation in the A(BD) and A(CD) trial types, viz, A1(B1D1), A4(B4D4), A9(C9D9), and A14(C14D14).

Visual inspection of Figures 2 and 3 suggests that there were more selections of the stronger relations by subjects who classified and rated first (Figure 3) than there were by those who matched to sample first (Figure 2). Indeed, the ratios of correct predictions (out of 240) were higher for the former ($t = 3.34$; $df = 35$; $p = .002$). Order of presentation of the two experiences (classifying/rating and matching to sample) may be responsible.

EXPERIMENT 4

Subjects in the first three experiments were exposed to the 80 stimuli in Table 1 in the classification and strength ratings and in the matching to sample. To this point, then, the generality of the results and the validity of the analysis are restricted to the stimuli on which the first three experiments were conducted. In the present experiment a new set of stimuli was constructed to determine whether the results of Experiments 1 to 3 could be replicated with different stimuli.

Moreover, in the prior experiments it was presumed (but not evaluated) that a continuum of similarity existed across the types of stimulus pairs, roughly described as running from A–A, A–B, A–C, to B–C, and from A–D, B–D, to C–D in which A–A stimuli were most similar (indeed, they were identical) and C–D stimuli were the least similar. This presumption was evaluated in the present experiment by having subjects rate the stimulus pairs on the basis of similarity.

METHOD

Subjects

College students were selected as they were in the prior experiments. Given the number of stimulus pairs to rate, five different groups of subjects were recruited (total = 265). The first three groups ($N = 150$) classified the stimulus pairs (taxonomic or thematic) and rated the strength of the relation portrayed in each pair as in Experiment 1. The last two groups ($n = 115$) rated the similarity of the stimuli comprising each pair. The average age of the subjects was 21.0 yrs.

Procedures

Stimuli. An additional 184 stimuli were selected from Corel Magic 200,000 and arrayed in pairs as in Experiment 1. Three hundred twenty-two slides were produced that contained these pairs, ordered by the number of the to-be-sample stimulus (i.e., 1A–46A) and second by the putative relation to the sample of the second stimulus (A–A, A–B, A–C, A–D, B–C, B–D, C–D). Table 8 lists all of these picture stimuli.

Classification and strength rating. The subjects in the first three groups experienced the procedures of Experiment 1 with the new stimulus set. Additionally, their procedures were different from Experiment 1 in two other ways. First, rather than rate all 322 stimulus pairs, the set was divided into three subsets of stimulus pairs of 107–108 each and each of the groups of subjects rated one of these subsets. Second, for these subjects, the 1–5 strength rating scale was projected via transparency and overhead projector on a screen to the left of the screen on which the pairs of stimuli were presented. The stimulus pairs were alone on the remaining screen except for the number of the slide to identify the correct line to mark on the Scantron.

Similarity rating. The subjects in the last two groups experienced the procedures of Experiment 1 with the new stimulus set but their instructions differed. For these groups, on another transparency, the overhead projector portrayed a 5-item Likert scale, but it stated: 5 = so similar they are identical; 4 = very similar; 3 = somewhat similar; 2 = a little bit similar; 1 = not similar. Subjects in these two groups were

Table 8. The 184 stimuli used in Experiments 4 and 5.

Sample No.	Sample Stimulus (A)	Basic Taxon (B)	Superordinate Taxon (C)	Theme (D)
1	Butterfly 1	Butterfly 2	Bee 1	Net
2	Lock 1	Lock 2	Chain lock	Keys
3	Hat 1	Hat 2	Hat 3	Boots
4	Sun glasses 1	Sun glasses 2		Magnifying glass
5	Watch 1	Watch 2	Sundial	Hand
6	Anchor 1	Anchor 2	Hook	Ship
7	Gas pump 1	Gas pump 2	Air pump	Car 1
8	TV 1	TV 2	VCR	Projection screen
9	Key 1	Key 2	Key 3	Lock 1
10	Hat 2	Hat 4	Hat 5	Man 1
11	Stereo 1	Stereo 2	Record player	Tape 3
12	Airplane 1	Airplane 2	Helicopter 3	Pilot
13	Dart 1	Dart 2	Arrow	Bullseye
14	Race car 1	Car 2	Truck	Race flags
15	Screw 1	Screw 2	Nail 1	Bolt
16	Saw	Chain saw	Hammer 3	Wood
17	8 ball	12 ball	Soccer ball 2	Pool table
18	Tape 1	Tape 2	Compact disc	Walkman
19	Dragon 1	Dragon 2	Unicorn	Knight
20	Bath tub 1	Bath tub 2	Sink	Towels
21	Squid 1	Squid 2	Stingray	Ocean
22	Caveman 1	Caveman 2	Mammoth 1	Hut
23	Cigarette 1	Cigarette 2	Pipe	Lighter
24	Baby 1	Baby 2	Bottle	Man 1
25	Hammer 1	Hammer 2	Sledge hammer	Nail 2
26	Helicopter 1	Helicopter 2	Airplane 4	Parachute
27	Soccer ball 1	Tennis ball	Football	Foot
28	Flashlight 1	Flashlight 2	Lighthouse	Battery
29	Car 1	Car 4	Bus	Gas can
30	Banana	Orange	Hamburger	Monkey 3
31	Squirrel	Mouse 2	Tiger	Peanut 1
32	Gun 1	Gun 2	Knife	Soldier
33	Fish 1	Fish 2	Shark	Fishing Fly
34	Horse 1	Horse 2	Donkey	Saddle
35	Elephant 1	Mammoth 1	Giraffe	Peanut 2
36	Monkey 1	Monkey 2	Ape	Tree
37	Cat 1	Cat 2	Lion	Yarn
38	Owl 1	Owl 2	Eagle	Feather
39	Telephone 1	Telephone 2	Cellular telephone	Ear
40	Computer 1	Computer 2	Laptop computer	Disk
41	Cheese 1	Cheese 2	Milk	Mouse 1
42	Grapes 1	Grapes 2	Cherry	Wine
43	Chicken 1	Chicken 2	Goose	Eggs
44	Carrot 1	Carrot 2	Broccoli	Rabbit
45	Spider 1	Spider 2	Scorpion	Web
46	Bee 2	Wasp	Fly 1	Beehive

Table 9. Classification Frequencies for the 322 Stimulus Pairs (Experiment 4).

Pair No.	A		B		C		D	
	Tax	Theme	Tax	Theme	Tax	Theme	Tax	Theme
1	28	2	25	5	46	31	2	28
2	77	2	70	9	17	12	0	38
3	78	1	27	3	17	13	3	27
4	27	2	27	3	21	54	1	28
5	38	1	26	4	46	34	4	75
6	78	0	76	2	16	14	1	29
7	29	1	25	5	11	17	0	38
8	38	0	36	1	11	19	3	26
9	80	0	35	4	75	4	2	37
10	37	1	74	6	34	4	1	79
11	39	0	37	2	16	13	29	1
12	39	0	76	4	30	8	5	72
13	80	0	76	4	22	8	6	74
14	77	2	24	5	52	26	0	39
15	29	1	76	1	35	44	6	24
16	79	0	34	5	14	15	0	29
17	29	1	26	4	25	14	7	32
18	79	1	74	5	36	42	3	36
19	72	3	79	1	9	29	3	76
20	79	0	28	11	17	20	7	31
21	29	1	27	3	14	15	2	28
22	29	1	61	18	3	35	1	28
23	29	1	27	3	29	9	3	35
24	80	0	65	14	40	39	1	37
25	38	1	37	1	69	11	0	39
26	29	1	35	3	47	33	8	71
27	77	1	31	8	20	9	1	78
28	80	0	34	2	24	56	3	36
29	28	2	77	1	14	16	9	71
30	29	1	15	15	10	20	2	28
31	78	1	70	7	6	24	1	78
32	75	2	26	4	6	24	2	77
33	77	1	33	6	31	6	3	27
34	79	0	29	1	15	15	10	69
35	29	1	60	18	28	52	0	37
36	28	2	26	4	34	4	2	78
37	78	1	73	7	51	29	1	38
38	29	1	73	6	64	13	6	74
39	39	0	35	4	37	2	0	38
40	79	1	28	2	26	4	3	27
41	80	0	25	5	8	22	1	37
42	39	0	37	2	9	28	5	34
43	28	2	28	2	32	7	2	28
44	76	2	38	1	18	12	3	27
45	28	0	28	1	9	21	3	26
46	78	1	68	11	17	13	12	67
Sums	2,480	42	2,037	233	1,211	940	167	2,032
Percent	98.3	1.7	89.7	10.3	56.3	43.7	7.6	92.4

instructed that: "We want you to judge the similarity of pairs of objects such as you see on the screen to your right. These are two dogs; since they are identical, you would mark a 5 on your Scantron form. A 5 means what the screen to your left says. Please don't mark the Scantron form until I say so. Here are two different dogs; they are not identical but they are very similar, so you would mark a 4 on your Scantron; now here are a dog and a cat; they are similar but not so much; so they might get a rating of 3 or a rating of 2; finally, here are a dog and a piece of meat. They are not similar at all, so they would get a rating of 1. You will judge the similarity of quite a few of these pairs. In order to know where you are, there are numbers on the left hand lower corner of each slide (pointed to). That number tells you what line to be marking on the Scantron. We'll occasionally announce the line also to keep everyone together. The slides will come up automatically. Any questions? Start on Answer 1. Let's begin."

The 322 stimulus pairs were divided in two and one subset of 161 stimulus pairs was presented to each of the last two groups. For all the groups, slides were presented at the rate of one every 7.5 s. As in the prior experiments, the presentation order of the pictures was randomized with restrictions that the picture on the left of the pair could not appear twice consecutively and the type of exemplified relation could occur no more than twice consecutively.

Results

Classification. The classification results are shown in Table 9. There was 98.3% judgment in favor of the taxonomic class given A-A pairs, 89.7% judgment in favor of the taxonomic class given the A-B pairs, 56.3% judgment in favor of the taxonomic class given the A-C pairs, and 92.4% judgment in favor of the thematic class given A-D pairs. Of the 46 A-C pairs, 16 were classed by a majority as thematic (i.e., A4-C4; A15-C15; A16-C16; A18-C18; A19-C19; A20-C20; A21-C21; A22-C22; A28-C28; A29-C29; A30-C30; A31-C31; A35-C35; A41-C41; A42-C42; A45-C45) and the subjects evenly divided on A34-C34. A11-D11 was the only A-D pair judged taxonomic by the majority.

Relation strength. Table 10 shows the mean relation strength for each of the stimulus pairs in the seven categories as judged by the sub-

jects. With this set of stimuli, the mean strength rating of the A-A pairs was less than 5 (Mean of Means = 4.32; $S^D = 0.09$). A-B pairs were judged lower (Mean of Means = 3.82; $S^D = 0.31$) and A-C pairs were judged lower yet (Mean of Means = 3.33; $S^D = 0.27$) with B-C pairs about the same (Mean of Means = 3.32; $S^D = 0.32$). Thematic A-D pairs were rated somewhat stronger than the superordinate pairs (i.e., A-C; B-C) (Mean of Means = 3.68; $S^D = 0.32$). Thematic B-D pairs rated about the same as thematic A-D pairs (Mean of means = 3.53; $S^D = 0.45$), while the lowest strength ratings went to the thematic C-D pairs (Mean of means = 2.95; $S^D = 0.56$). The additional strength ratings in this experiment that are shown in Table 10 (i.e., B-C, B-D, and C-D ratings) are not used in eventual matching-to-sample predictions because B and C stimuli were never used as samples in that task. Different from Experiments 1 through 3, strength ratings on these pairs were necessary in this experiment to compare with similarity ratings on the same pairs. (See below.)

Because subjects experienced different pair types and there were different numbers of subjects in each of the three groups, three, one-way, within-subjects, analyses of variance of the relation strength ratings were conducted with each type of stimulus pair as factors. For each factor, there was a statistically significant effect of type of stimulus pair on the mean relation strength ratings. For Group 1, Wilks' $\Lambda = .611$; $F(6, 24) = 2.55$; $p < .05$; $\eta^2 = .39$. Follow up, pair-wise comparisons showed that of the 21 post hoc contrasts, mean relation strength differed statistically between 17 pairs of means ($p < .05$), and did not differ statistically on 4 contrasts, i.e., A-B v A-D ($p < .80$), A-C v B-D ($p < .09$), A-C v B-C ($p < .50$), B-C v B-D ($p < .06$). For Group 2, Wilks' $\lambda = .44$; $F(6, 74) = 15.93$; $p < .001$; $\eta^2 = .56$. Follow up, pair-wise comparisons showed that of the 21 post hoc contrasts, the mean relation strength ratings differed statistically between 20 pairs of means ($p < .05$) and did not differ statistically for one, A-D v B-D ($p = .17$). For Group 3, Wilks' $\lambda = .53$; $F(6, 33) = 4.95$; $p = .001$; $\eta^2 = .47$. Follow-up, pair-wise comparisons showed that of the 21 post hoc contrasts, mean relation strength ratings differed statistically between 20 pairs of means ($p < .05$) and did not differ statistically for one, A-C v B-C ($p = .06$).

Table 10. Mean Strength Ratings for Stimulus Pairs in Experiment 4.

Pair No.	AA	AB	AC	AD	BC	BD	CD
1	4.23	3.70	3.06	3.47	2.90	3.47	3.33
2	4.38	3.75	3.37	3.87	3.44	3.23	3.13
3	4.38	3.60	3.23	3.90	3.15	3.66	2.22
4	4.20	3.67	2.81	3.31	3.04	3.79	3.11
5	4.31	3.55	3.42	3.09	2.97	2.43	2.41
6	4.27	4.06	3.10	3.83	3.07	3.65	3.51
7	4.33	3.77	3.03	3.47	3.07	3.13	3.10
8	4.23	3.56	3.37	3.50	3.36	4.03	3.18
9	4.47	4.10	3.91	4.00	3.90	3.88	3.54
10	4.21	3.95	3.54	2.84	3.72	2.87	2.04
11	4.18	3.87	3.10	3.44	3.17	3.90	3.46
12	4.38	4.04	3.69	3.79	3.56	3.97	3.69
13	4.47	4.10	3.37	4.15	3.59	4.07	3.40
14	4.40	3.30	3.43	3.97	3.36	2.72	2.54
15	4.33	4.31	3.27	3.87	3.20	3.80	2.90
16	4.40	3.82	3.03	3.30	2.93	3.20	3.23
17	4.30	3.77	3.26	4.00	3.30	3.97	2.74
18	4.40	4.10	3.63	4.00	3.20	3.50	2.97
19	4.19	4.07	3.10	3.47	3.15	3.45	2.70
20	4.46	4.18	3.46	3.73	3.40	3.92	3.74
21	4.33	3.47	3.37	3.40	3.30	3.20	3.21
22	4.27	3.72	3.26	3.40	3.44	3.24	3.03
23	4.33	3.97	3.87	3.87	3.85	3.90	3.53
24	4.41	3.60	3.25	4.05	3.41	3.33	1.79
25	4.23	4.23	3.83	4.13	3.90	4.00	3.74
26	4.30	4.23	3.69	2.79	3.52	3.21	3.01
27	4.19	3.64	3.33	3.85	3.08	2.90	3.64
28	4.44	4.26	2.99	3.74	2.79	3.71	2.05
29	4.30	3.50	3.27	3.85	2.97	3.71	3.00
30	4.33	3.37	3.17	3.93	3.23	2.09	1.69
31	4.41	3.08	2.67	3.37	2.63	3.15	2.13
32	4.35	3.8	3.17	3.67	3.51	3.00	2.97
33	4.10	3.46	3.34	3.63	3.33	3.27	2.53
34	4.41	3.30	3.23	3.45	3.54	3.55	3.20
35	4.30	3.59	2.85	3.82	2.49	3.68	2.46
36	4.30	4.11	3.72	3.35	3.87	3.74	3.06
37	4.33	4.08	3.32	3.33	3.41	4.18	1.86
38	4.33	3.85	3.30	3.68	3.41	3.67	3.73
39	4.36	3.95	3.87	3.90	3.83	3.95	3.62
40	4.46	3.87	3.43	3.57	3.79	3.83	3.62
41	4.41	3.97	3.40	3.92	3.40	3.56	2.07
42	4.33	4.15	3.18	4.08	3.21	3.64	2.92
43	4.30	3.77	3.51	3.90	3.28	4.08	2.77
44	4.33	4.23	3.27	3.83	3.27	3.67	3.23
45	4.20	3.10	3.26	3.80	3.47	3.90	2.77
46	4.32	3.92	3.23	3.86	3.37	3.95	2.99
Means	4.32	3.82	3.33	3.68	3.32	3.53	2.95

Figure 4. Mean similarity ratings by the subjects of Experiment 4 of the 322 stimulus pairs divided into seven sets of pair types (i.e., A–A, A–B, etc.). Pair types A–A, A–B, and A–C relate taxonomic stimuli, while pair types A–D, B–D, and C–D relate a taxonomic and a thematic stimulus.

Similarity.

The results of the similarity ratings are shown in Figure 4. Each panel shows the mean similarity ratings for each of the 46 pairs of stimuli

that comprised each of the seven pair types (i.e., A–A, A–B, etc.). An orderly relation can be observed across the pair types with the highest similarity ratings given to the A–A pairs and the lowest given to the C–D pairs.

Figure 5. Mean similarity ratings within pair types by the subjects of Experiment 4 for all of the pair types (i.e., A–A, A–B, etc.). Pair types A–A, A–B, and A–C relate taxonomic stimuli, while pair types A–D, B–D, and C–D relate a taxonomic and a thematic stimulus.

The relation of similarity to pair type is also orderly within the two general pair types (i.e., taxonomic and thematic) as can be seen in Figure 5 which shows the mean similarity rating for each of the panels of Figure 4. Pair types A–A, A–B, A–C, and B–C each relate two taxonomic stimuli, while pair types A–D, B–D, and C–D each relate a taxonomic and a thematic stimulus. Similarity ratings decreased from A–A to B–C and from A–D to C–D.

Because subjects experienced different pair types and there were different numbers of subjects in each of the two groups doing similarity ratings, two, one-way, within-subjects, analyses of variance of similarity ratings were conducted with each type of stimulus pair as factors. For each group, there was a statistically significant effect of type of stimulus pair on the mean similarity rating. For Group 1, Wilks' $\lambda = .02$; $F = 426.84$; $df = 6, 40$; $p < .001$; $\eta^2 = .98$. For Group 1, follow up, pair-wise comparisons showed that mean similarity differed statistically between all pairs of means ($p < .001$), except A–D v B–D ($p = 0.10$). For Group 2, Wilks' $\lambda = .211$; $F = 39.17$; $df =$

6, 63; $p < .001$; $\eta^2 = 0.79$. For Group 2, follow-up, pair-wise comparisons showed that mean similarity differed between all pairs of means ($p < .001$), except A–C v B–C ($p = .12$).

Figure 6 depicts the bivariate correlation between the mean strength ratings and the mean similarity ratings for each of the 322 stimulus pairs. There is a strong correlation ($r = 0.75$; $df = 321$; $p < .01$), suggesting that strength and similarity account for about 55% of the variation between these two variables. Next, the pairs were split into those that related two taxonomic stimuli (A–B, A–C, B–C) and those that related one taxonomic and one thematic stimulus (A–D, B–D, C–D). Those that related two taxonomic stimuli correlated even more strongly ($r = .81$; $df = 160$; $p < .01$), while those that related one taxonomic and one thematic stimulus correlated less strongly ($r = .55$; $df = 160$; $p < .01$). Thus, similarity and relation strength appear to account for 65% of the variance in those pair types that relate taxonomies, while relation strength and similarity account for only 30% of the variation in the pair types that are thematically related.

Figure 6. Scatterplot of the relation ($r = .75$) between similarity and strength ratings in Experiment 4.

Discussion

In this experiment a new set of stimuli was composed by different experimenters, essentially without regard for (but with knowledge of the framework of) the set of stimuli used in Experiments 1 through 3. With this new set of stimuli, the results for the classification and strength ratings replicate those of the prior experiments. From A-A pairs, judged mostly taxonomic, taxonomic judgments decreased until on A-C pairs, fully 35% were classed by a majority of subjects as thematic—the size of this outcome unforeseen—while almost all A-D pairs were classed thematic. Subjects' judgments of relation strength rated the A-A pairs most strongly and the strength ratings decreased across the pair types in orderly fashion until they increased again for the A-D thematic ratings.

The similarity ratings paralleled the strength ratings with A-A pairs rated most similar. Similarity then decreased from basic taxon through superordinate taxon. The taxonomic pairs (A-A, A-B, A-C, B-C) were rated higher on simi-

larity than the thematic pairs (A-D, B-D, C-D).

The higher correlation between similarity and strength ratings for the taxonomic pair types suggests that similarity could be more of a controlling factor on those pair types than on the thematic pair types.

EXPERIMENT 5

In this experiment we wanted to systematically replicate Experiment 2 with the new set of picture pairs used in Experiment 4 by placing those picture pairs into a conditional discrimination task, i.e., matching to sample, of different trial types with common samples. Again we wanted to determine whether there would be systematic selections of identical, basic, or superordinate, taxonomically related comparisons and/or thematically related comparisons in the various trial types; and, if the strength ratings generated in Experiment 4 related to comparison selection in this experiment.

METHOD

Subjects

As before, subjects ($N = 63$) were volunteers from beginning undergraduate classes at Utah State University. Inadvertently, subjects' ages were not coded. Subjects were again offered extra credit for their participation in the experiment.

Stimuli

Five hundred fifty-two slides were composed from the stimuli described in Table 8 as they were in Experiment 2.

Procedure

The procedure was the same as Experiment 2 except that the subjects were subdivided into three groups, all of whom matched-to-sample 250 configurations of the 552—the scoring limit of the Scantron form. Group 1 matched configuration numbers 1–250, Group 2 numbers 402–552, and Group 3 numbers 200–450.

Results

Table 11 shows the collected responses of the subjects to each trial type (columns) for each sample (rows). The table shows that for trial types A(AB), A(AC), and A(AD) where the sample stimulus could be matched with itself, that is, A was a sample and the same A was a comparison, subjects most frequently chose A. Whenever an arbitrary match was required—on trial types A(BC), A(BD), and A(CD)—subjects chose B when B and C were comparisons; chose B when B and D were comparisons; and chose C less frequently than D when C and D were comparisons (see column totals).

Table 11 also shows the results of relating comparison selection with the mean strength of relation ratings that were produced in Experiment 4. Again, the more strongly rated relation was used to predict comparison selection in the matching-to-sample task. The cells in Table 11 contain asterisks for 45 of the 46 configurations on trial type A(AB), 46 of 46 configurations on trial types A(AC) and A(AD); 44 of 46 configurations for trial type A(BC);

35 of 46 configurations for trial type A(BD); and 34 of 46 configurations for trial type A(CD).

Discussion

Collectively, the stronger stimulus relation was chosen on 97.8% of the configurations on the matching to sample when the trial type was A(AB); on 100% when the trial types were A(AC) and A(AD); on 95.6% of the configurations when the trial type was A(BC); on 76.1% when the trial type was A(BD); and on 73.9% when the trial type was A(CD). In two configurations, A14(B14C14) and A45(B45C45), B was selected though A–B was the more weakly rated relation. Uniquely, the A25–A25 and A25–B25 relations tied in strength ratings (Table 10), but all but five responses were selections of A25–A25 on the A25(A25B25) configuration. On the 11 incorrect predictions for the A(BD) trial type, eight were selections of A–B when A–D was the more strongly rated relation, two were selections of A–D when A–B was the more strongly rated relation, and one was an equal selection of A–B and A–D when A–B was the (slightly) more strongly rated relation. Overall, the more strongly rated of the two relations was selected on 250 of the 276 configurations (90.6%). When the relation strength rating is not predictive, similarity between sample and comparison may prevail.

On configurations of the A(CD) trial type where similarity between sample and comparison is less an issue, 41 of the 46 A–D relations were rated stronger than their comparable A–C relations. The differential strength analysis predicted comparison selection on 29 of these 41 correctly (70.7%). Four of the A–C relations were rated stronger than their comparable A–D relations. All four were predicted correctly (100%). One set of relations (A23–C23 and A23–D23) was tied in strength ratings (Table 10). There were 63 total responses to the A23(C23D23), configuration so that given the odd number, a strict tie was not possible. The A23–D23 relation was selected 32 times and the A23–C23 relation was selected 31 times (Table 11), as close to a tie as was possible. To reiterate, in the A(CD) trial type, if the strength relation only reflected similarity between sample and comparison, there should have been no correct A–D predictions.

Table 11. Frequencies of Selection of Comparisons (Experiment 5).

Sample (A)	CO:AB	CO:AC	CO:AD	CO:BC	CO:BD	CO:CD
A1	A1=63* B1=0	A1=59* C1=4	A1=51* D1=7	B1=44* C1=6	B1=19* D1=17	C1=34 D1=45*
A2	A2=77* B2=2	A2=57* C2=6	A2=45* D2=18	B2=68* C2=11	B2=40 D2=23	C2=25 D2=38*
A3	A3=74* B3=5	A3=71* C3=3	A3=45* D3=18	B3=44* C3=6	B3=17 D3=35	C3=8 D3=21*
A4	A4=44* B4=3	A4=55* C4=13	A4=35* D4=12	B4=54* C4=7	B4=60* D4=14	C4=34 D4=28
A5	A5=49* B5=1	A5=56* C5=2	A5=55* D5=8	B5=70* C5=4	B5=32* D5=15	C5=34* D5=29
A6	A6=57* B6=6	A6=77* C6=2	A6=27* D6=9	B6=95* C6=2	B6=39* D6=11	C6=18 D6=40*
A7	A7=61* B7=2	A7=65* C7=3	A7=48* D7=20	B7=56* C7=5	B7=51* D7=17	C7=24 D7=50*
A8	A8=57* B8=4	A8=76* C8=3	A8=46* D8=6	B8=89* C8=8	B8=61* D8=30	C8=24 D8=55*
A9	A9=72* B9=2	A9=44* C9=3	A9=36* D9=11	B9=43* C9=2	B9=39* D9=24	C9=33 D9=25
A10	A10=58* B10=1	A10=59* C10=4	A10=38* D10=7	B10=51* C10=1	B10=54* D10=9	C10=40* D10=33
A11	A11=53* B11=5	A11=51* C11=1	A11=71* D11=8	B11=49* C11=1	B11=66* D11=13	C11=23 D11=38*
A12	A12=100* B12=8	A12=61* C12=2	A12=37* D12=13	B12=30* C12=4	B12=60* D12=14	C12=23 D12=24*
A13	A13=31* B13=3	A13=34* C13=0	A13=34* D13=8	B13=45* C13=7	B13=17 D13=19*	C13=22 D13=51*
A14	A14=35* B14=1	A14=29* C14=3	A14=36* D14=9	B14=44 C14=19	B14=10 D14=37*	C14=14 D14=44*
A15	A15=30* B15=4	A15=34* C15=2	A15=23* D15=6	B15=32* C15=4	B15=34* D15=13	C15=15 D15=58*
A16	A16=25 B16=4	A16=94* C16=9	A16=76* D16=16	B16=35* C16=10	B16=38* D16=9	C16=8 D16=10*
A17	A17=70* B17=4	A17=59* C17=4	A17=23* D17=11	B17=56* C17=5	B17=31 D17=16	C17=17 D17=62*

(Continued, p. 82)

Table 11. Frequencies of Selection of Comparisons (*continued*).

Sample (A)	CO:AB	CO:AC	CO:AD	CO:BC	CO:BD	CO:CD
A18	A18=57* B18=6	A18=50* C18=2	A18=12* D18=6	B18=49* C18=3	B18=26 D18=26	C18=12 D18=46*
A19	A19=54* B19=4	A19=50* C19=4	A19=51* D19=12	B19=39* C19=6	B19=45* D19=13	C19=9 D19=9
A20	A20=86* B20=4	A20=59* C20=4	A20=33* D20=12	B20=59* C20=15	B20=46* D20=15	C20=36 D20=27
A21	A21=29* B21=3	A21=43* C21=4	A21=50* D21=7	B21=34* C21=2	B21=53* D21=12	C21=68 D21=35
A22	A22=72* B22=7	A22=106* C22=13	A22=69* D22=23	B22=35* C22=12	B22=33 D22=35	C22=18 D22=45*
A23	A23=39* B23=8	A23=45* C23=2	A23=38* D23=9	B23=52* C23=10	B23=40* D23=17	C23=31* D23=32*
A24	A24=53* B24=5	A24=57* C24=5	A24=61* D24=13	B24=25* C24=7	B24=8 D24=20	C24=10 D24=53*
A25	A25=53 B25=5	A25=28* C25=6	A25=60* D25=14	B25=47* C25=5	B25=36* D25=22	C25=15 D25=37*
A26	A26=27* B26=2	A26=45* C26=5	A26=18* D26=0	B26=43* C26=6	B26=16* D26=2	C26=52* D26=11
A27	A27=64* B27=4	A27=58* C27=5	A27=60* D27=13	B27=29* C27=23	B27=4 D27=14*	C27=25 D27=33*
A28	A28=69* B28=5	A28=58* C28=5	A28=27* D28=9	B28=40* C28=4	B28=38* D28=14	C28=28 D28=35*
A29	A29=61* B29=7	A29=64* C29=3	A29=53* D29=8	B29=82* C29=5	B29=40 D29=22	C29=17 D29=17
A30	A30=56* B30=7	A30=46* C30=1	A30=49* D30=14	B30=33* C30=1	B30=16 D30=18*	C30=5 D30=42*
A31	A31=63* B31=5	A31=74* C31=5	A31=13* D31=3	B31=44* C31=3	B31=14 D31=22*	C31=8 D31=39*
A32	A32=58* B32=3	A32=72* C32=2	A32=38* D32=9	B32=53* C32=5	B32=45* D32=18	C32=13 D32=21*
A33	A33=47* B33=3	A33=47* C33=5	A33=64* D33=10	B33=39* C33=23	B33=67 D33=23	C33=50 D33=29

(Continued, p. 83)

Table 11. Frequencies of Selection of Comparisons (*continued*).

Sample (A)	CO:AB	CO:AC	CO:AD	CO:BC	CO:BD	CO:CD
A34	A34=68* B34=6	A34=59* C34=2	A34=39* D34=8	B34=49* C34=3	B34=24 D34=12	C34=46 D34=44
A35	A35=71* B35=3	A35=70* C35=4	A35=54* D35=9	B35=51* C35=7	B35=20 D35=12	C35=20 D35=30*
A36	A36=46* B36=6	A36=33* C36=1	A36=38* D36=9	B36=26* C36=3	B36=30* D36=17	C36=28* D36=19
A37	A37=43* B37=2	A37=29* C37=5	A37=52* D37=6	B37=23* C37=13	B37=51* D37=23	C37=38 D37=28
A38	A38=87* B38=9	A38=102* C38=6	A38=47* D38=5	B38=48* C38=2	B38=30* D38=4	C38=42 D38=16
A39	A39=59* B39=4	A39=27* C39=2	A39=57* D39=6	B39=64* C39=16	B39=52* D39=9	C39=38 D39=9
A40	A40=50* B40=2	A40=57* C40=6	A40=26* D40=10	B40=55* C40=6	B40=71* D40=26	C40=20 D40=16
A41	A41=59* B41=9	A41=59* C41=4	A41=65* D41=14	B41=44* C41=3	B41=8 D41=10	C41=21 D41=42*
A42	A42=31* B42=3	A42=88* C42=2	A42=52* D42=11	B42=59* C42=4	B42=29* D42=18	C42=13 D42=34*
A43	A43=57* B43=6	A43=45* C43=0	A43=23* D43=11	B43=33* C43=3	B43=38 D43=23	C43=15 D43=30*
A44	A44=53* B44=5	A44=0** C44=0	A44=37* D44=6	B44=46* C44=6	B44=32* D44=15	C44=23 D44=39*
A45	A45=56* B45=4	A45=62* C45=1	A45=50* D45=13	B45=60 C45=8	B45=33 D45=30	C45=1 D45=35*
A46	A46=61* B46=2	A46=107* C46=5	A46=52* D46=11	B46=18* C46=0	B46=44* D46=29	C46=11 D46=22*
Totals	A=2,585 (93.0%) B=194 (7.0%)	A=2,621 (93.9%) C=171 (6.1%)	A=2,014 (80.8%) D=478 (19.2%)	B=2,184 (87.7%) C=306 (12.3%)	B=1,666 (66.6%) D=837 (33.4%)	C=1,110 (42.1%) D=1,526 (57.9%)
% Hits	97.8	100.0	100.0	95.6	76.1	73.9

*Matching-to-sample outcome predicted by mean strength of relation rating from Table 10. The 26 cells of the table that are outlined are not predicted by the results of Table 10.

**A44(A44C44) and A44(C44A44) were inadvertently omitted from the presentation.

As with the earlier set of stimuli, subjects appeared to match on the basis of perceptual similarity of sample and comparison on trial types where that was possible. This is inferred from the A(AX) trial types. And it can also be inferred from the arbitrary matching-to-sample trial types as well. As B comparisons were, indeed, rated more perceptually similar to A comparisons than C comparisons, subjects chose B comparisons to match with A in the A(BC) trial type. As B comparisons were rated more perceptually similar to A comparisons than D comparisons were, subjects chose B comparisons to match with A in the A(BD) trial type. Finally, where there was little or no similarity between A samples and C or D comparisons, selections were evenly distributed between C comparisons and D comparisons in the A(CD) trial type.

GENERAL DISCUSSION

Developmental Psychology and These Studies

In two-choice matching to sample utilizing natural stimuli that are related by theme or taxon, college students frequently select the stimulus relation that is—by themselves or by others like them—the most strongly rated. Relation strength ratings may relate to comparison selection in this matching-to-sample task only in the absence of modeling or reinforcement, because the latter two variables exert considerable control when present (Bauer & Mandler, 1989; Osborne & Koppel, 2001; Smiley & Brown, 1979). In the present studies, there was no formal modeling, and with the exception of the presentation of a subsequent trial, there were no systematic outcomes that were a result of responding. To this point it is interesting to note that in the Smiley & Brown (1979) study, the model was observed in a first session, and the model's behavior was influential in providing a context for subject behavior as either thematic or taxonomic. In the next session, in the absence of the model, subjects tended to revert to the type of selections they had made prior to observing the model in the prior session (Smiley & Brown, 1979). Explicit or implicit ratings of relation strength may be valid only in the absence of (programmed) reinforcement and modeling.

As age was not manipulated, generalization

of the utility of this ratings strength analysis to other ages is unwise until such time as it is shown to apply there as well. However, Osborne and Calhoun (1998) showed that young children (29–73 mos), across a number of experiments using a similar task—2—choice matching to sample in which the configurations were primarily of the A(CD) type—tended to respond similarly across experiments to individual configurations. That is, on some configurations most of the children selected taxonomically, on others most selected thematically, and on still others they were, collectively, indifferent (see also Bauer & Mandler, 1989). These results occurred independently of the kind of taxon, instructions, general praise for responding or its absence, or the presence of a third unrelated comparison in the matching-to-sample task. However, Osborne and Calhoun (1998) did not ask their subjects to rate the stimulus relations in the matching-to-sample task.

The present results suggest one reason why taxonomic or thematic selection prevails on a given configuration or trial type. Subjects are not necessarily more taxonomically or thematically oriented. Rather, there are configurations where they are more likely to choose thematically and configurations where they are more likely to choose taxonomically (Osborne & Calhoun, 1998). The results of the present studies suggest that this differential likelihood relates to the different strength ratings of the stimulus relations that comprise each configuration. That is, each matching-to-sample configuration comprises its own context for each subject given the subjects' histories with the stimulus relations therein.

Experimental Analysis of Human Behavior and These Studies

The present experiments are hybrids in the sense that they relate to what has been an issue for developmental psychologists, but they do so from the framework of behavior analysis and have implications for behavior analysis as well. To developmental psychologists issues surrounding taxonomic and thematic relational responding may be thought to be settled by demonstrations that such responding relates collectively to reinforcement and modeling (Bauer & Mandler, 1989; Smiley & Brown, 1979) and verbal justification by subjects as to

why they selected a particular comparison in a particular configuration (e.g., Baldwin, 1992). To the present authors, the question of relational responding in this conditional discrimination task is not settled, until—as with Watson's goal—knowing the effective stimulus (relation), we can predict the response reasonably well. The ratings strength analysis attempts a closer approach to that goal.

Further, the experiments herein constitute functionally direct or systematic replications of one another—indeed, replications in which the results on several levels were nearly identical (e.g., relation classification; strength ratings; correct predictions). These replications suggest the irrelevance of subject history, although subject history was not directly manipulated (except for immediate history in Experiment 3) and the subjects in each experiment were from the same age and cultural population. Indeed, one of the replications suggests the generality of the results (and the analysis) across different sets of stimuli.

Natural and other stimuli. The use of natural stimuli in experimental analysis has solid precedent. Such stimuli are useful in the study of conceptual and verbal behavior and reasoning (e.g., Herrnstein, Loveland, & Cable, 1976; Premack, 1983). However, the tendency of most experimental analysts of human behavior is to avoid the use of natural stimuli since they are difficult to describe in the language of physics, and they have meaning to human subjects. Meaning may be thought to confound acquisition. In the acquisition of conditional discriminations, this has led to the use of arbitrary stimuli and symbols from other languages which are assumed to have no meaning to most westerners (e.g., Gatch & Osborne, 1989; Sidman & Tailby, 1982). Yet, the fact that natural stimuli have meaning to human subjects is critical to the examination of such relations as taxons and themes (see also, Bush, Sidman, & deRose, 1989). Moreover, it is difficult to see how taxonomic and thematic stimulus relations could be examined without resorting to the use of natural stimuli. In any case, acquisition was not a question in the present studies.

Multiple stimulus control topographies. With limited exceptions (Dube & McIlvane, 1996; Lowenkron, 1998), behavior analytic explanations of matching-to-sample appear—by omission—to consider that a single stimulus relation is functional on any trial. That is, responses

to A1–B1 in a configuration composed of A1(B1B2) are reinforced and behavior analysts consider primarily the development of the singular A1–B1 relation and ignore A1–B2 if possible (sometimes with peril, e.g., Carrigan & Sidman, 1992; Stromer & Osborne, 1982). Even when responses to A1–C1 are reinforced to set the stage for the emergence of transitive relations, A1–C2 is ignored where possible. Control by stimulus relations not specified by the experimenter suggests the existence of differing stimulus control topographies (McIlvane & Dube, 1992).

Most directly related to the present experiments are configurations such as A1(B1C1) after the foregoing training. That is, A1(B1C1) is a configuration with two functional stimulus relations with a somewhat equivalent relational history—the kinds of configurations used in the present studies. Responses to such configurations are necessarily forced choice and we say that one relation controlled on that particular trial even though both relations are potentially functional. Where it has been demonstrated that more than one relation exists, and selection of one or the other comparison occurs, the construct of selective attention is invoked (Ray, 1969), which describes, but does not explain, the resulting stimulus control where there is more than one possible controlling relation. In the natural environment, it seems reasonable to consider the existence of matching-to-sample configurations that consist of more than one functional stimulus relation, only one of which can be functional on an occasion because the subject can only make one response. Such configurations constitute a choice task and the variables known to be functional in choice tasks—for example, reinforcement history, quality of, and time to, reinforcement—should be functional (Dube & McIlvane, 1996). Again, such configurations suggest multiple stimulus control topographies.

Is it possible that the ratings strength analysis proposed herein might constitute a partial explanation of relation selection in the conditional discrimination task more broadly than in the present experiments? Such an analysis would not necessarily fly in the face of established explanations for responding in matching to sample (Carter & Werner, 1978; Cumming & Berryman, 1965) for non human subjects in which responding in the conditional discrimination is controlled by stimulus rela-

tions of the type if . . . then (i.e., if A1, then B1) and the conditional discrimination is thought to be composed of a successive discrimination (between samples) followed by a simultaneous discrimination (between comparisons), that is, is composed of two, three-term contingencies (Sidman, 2000).

In the present studies the subjects, by their ratings, the initial instructions, and their histories, were familiar with both of the relations in the matching-to-sample configurations. Where there are two functional stimulus relations that share a common stimulus it is reasonable to consider that the stimulus common to each must code for both of the stimuli to which it relates. As such, Lowenkron's (1998) analysis of selection in conditional discrimination tasks may relate to the present task. In his conception, selections in conditional discriminations that are relationally based are mediated by verbal responses such as tacts, self-echoics, self-duplicates, and autoclitics. The mediating verbal response identifies which kind of relation is exemplified by the two stimuli. In this conception, autoclitics that are produced by each of the pairs of stimuli in a matching-to-sample configuration cycle until one eventually leads to the selection response. Left in question is what ends the cycle. Could the differential strengths of the stimulus relations answer?

In the present set of studies two sets of operants were shown to be highly correlated. Subjects' behavioral ratings of the strengths of pairs of stimuli in one task related to the subsequent behaviors of the same and different subjects in a task that used the same pairs of stimuli in a conditional discrimination. The behaviors of rating the strengths of stimulus pairs were at different times both conceptually and temporally prior to the selections of relations in the subsequent matching-to-sample task and at other times came after matching to sample. That such responses are so highly correlated may suggest that the stimulus control topographies in each cohere. In other words, the subjects were essentially emitting different behaviors in the two tasks that were controlled by the same stimulus relations. Logically, this makes sense given the assumption that the same stimulus relations are operationally present in the two tasks and may be similarly functional therein.

Hypothetical constructs. It might be argued that the relationship between these operants is assumed and hypothetical, but that is not so.

Both strength ratings and comparison selections were publically emitted, were related statistically, and were found to be well correlated dependent variables. To functionally relate these operants would necessitate the manipulation of one or the other—presumably first the strength judgments of the subjects—and to determine that such manipulations related functionally to subjects' subsequent relation selections in the matching-to-sample task. Then, secondly, directed responding in the matching to sample, say by reinforcing selections of taxonomic or thematic comparisons, would presumably result in different strength ratings. The current studies contain several approximations to the former objective. First, an indirect manipulation of relation strength involved the construction of the stimulus sets in all of the experiments. No particular care was taken to design these stimulus sets other than to select recognizable stimuli and to give the stimulus relations the stimulus pairs were supposed to portray face validity. As such there was no design with respect to the construction of the stimulus pairs that led systematically to one relation being identified as stronger than any other, *a priori*. Relation strength ratings were free to vary in this context as were the differences between relations in terms of their rated strengths. As such, there was change of rated strengths within and across stimulus pair types—the manipulation of a parameter of an independent variable akin to exposing subjects to a random set of values of a particular parameter of an independent variable. Second, the use of a different set of stimuli (Experiments 4 and 5)—again constructed in the absence of prior design with respect to individual configurations—is a step closer to the systematic manipulation of an independent variable in these experiments. Such a change is equivalent to presenting a new random set of values of the same independent variable. Finally, the difference in the order of procedure in Experiment 3 constitutes another approximation to an independent variable. The subjects in Experiment 3 who had the experience of classifying and judging the strengths of the relations in their matching-to-sample configurations before they saw those configurations selected the stronger of the two stimulus relations more frequently than did the subjects in Experiment 3 who matched first and rated second. This outcome suggests that whatever stimulus con-

rol topography was produced by classifying and rating facilitated the matching to sample in a particular way.

Incomplete research designs. The absence of complete research designs in these experiments taken individually is not detrimental to the resulting inferences which attain generality via replications across experiments. Consider the absence of complete research designs in such demonstration experiments as those by Herrnstein et al. (1976) with respect to pigeon conceptual behavior and Premack (1983) with respect to reasoning by chimpanzees. In Herrnstein et al.'s (1976) experiments there was no direct manipulation of an independent variable. Generalization of outcome was achieved via systematic replication of the pigeons' conceptual behavior with more than one concept—akin to using more than one stimulus set in the present experiments. In the present experiments, when a new set of stimuli was constructed, the results were largely the same across experiments, suggesting the generality of the outcomes and the analysis that led to them. Moreover, having subjects classify and rate the strength of stimulus pairs prior to the inclusion of those same stimulus pairs in matching-to-sample configurations appeared to relate to subjects more frequently selecting the comparisons in the more strongly rated stimulus pairs than was the case for their peers who classified and rated strength after they did the matching to sample. In sum, from the present experiments, it is possible to infer that strength ratings constitute operants of similar stimulus control topography to the stimulus control topography involved in the subsequent selection of stimulus relations in the matching-to-sample task. Moreover, this conception is possible without considering ratings strength as a hypothetical construct. It is, rather, a derived dependent variable in its own right that is highly correlated with operants that follow.

Conditional discriminations: To be or not to be. And, finally, are the trial types used in the present experiments really conditional discriminations? In a typical conditional discrimination, the sample instructs which one of two or more concurrently presented comparison stimuli are discriminative for reinforcement. In the present experiments the sample could indicate that both comparison stimuli were discriminative for reinforcement—were there any reinforced trials, which there were not—be-

cause selection of either comparison in the present experiments was consistent with one or the other relations postulated by the experimenters. Under these circumstances one cannot call this task a conditional discrimination since the sample is not necessary to the selection of any one comparison. In other words, procedurally, samples herein do not appear to have an instructional function (Cumming & Berryman, 1965). However, this procedural absence does not necessarily translate to the absence of an instructional function of the sample in this task. Consider any conditional discrimination, where as above, in the configuration A1(B1B2) selection of B1 is reinforced; and in the configuration A1(C1C2) selection of C1 is reinforced. In each case A1 instructs the selection of a different comparison. In those two configurations A1's instructional function occurs as a result of differential reinforcement and the different comparisons shown with it on a particular trial. After such training we speak of relations between A1 and C1 and between A1 and B1. What then of a trial on which we present A1(B1C1) after such training? Does such a trial do away with the conditionality of the conditional relation? Procedurally, it would appear to. But such a conclusion neglects the subject's most recent history in which, across trials, samples are instructional. The select-one-relation-only condition continues on an A1(B1C1) trial, since after a selection of either comparison, the trial ends. The sequence of the experiment does not change. These conditions permit the inference of a generalized function (as instructor or selector of discriminations) of the sample, notwithstanding the absence of conditionality owing to both comparisons being consistent with previously reinforced relations.

There are several ways to circumvent the conditionality dilemma argumentatively, all of which involve demonstrations of the instructional function of the sample in this same task. In one, differential reinforcement of taxonomic or thematic selections gives the samples an instructional function (i.e., select taxonomic comparisons or select thematic comparisons). Osborne and Koppel (2001) demonstrated such a function. They reinforced college student selections of either taxonomic or thematic relations on a set of trial types of the A(CD) type (with A-C = superordinate taxonomic relations and A-D = thematic relations). After being

trained to criterion on one set of configurations, subjects were tested (in the absence of feedback) on a new set of configurations of the same type. The feedback contingencies were then reversed. Where subjects were given feedback for selecting taxonomically, they were given feedback for selecting thematically, and vice versa. After each of several reversals, subjects were exposed to new sets of configurations of the same type, in the absence of feedback. On the new sets of configurations, the samples instructed the kind of comparison selection that had been reinforced most recently.

Another way around the conditionality dilemma is to have each sample instruct differentially by placing it, too, under stimulus control, that is, under the control of a contextual stimulus (Sidman, 1986). Osborne and Koppel (2001) did this also in a related experiment using the same task. In this experiment, two additional stimuli (CS1 and CS2) were employed as contextual stimuli. On trials in which CS1 was present, responses to A–C relations were reinforced, while on trials in the presence of CS2, responses to A–D relations were reinforced. The instruction altering functions of CS1 and CS2 were noted by the differential selection of relations on configurations that the subject had never seen before in the absence of reinforcement. CS1 and CS2 functioned as generalized contextual stimuli instructing samples as to the related comparisons.

Finally, the function of the sample was indirectly shown by Osborne and Calhoun (1998) in this same task. In most of Osborne and Calhoun's five experiments, two-choice matching to sample was employed and the trial types were either A(BD) or A(CD). To counter speculation about the possibility of S-control (Stromer & Osborne, 1982), Osborne and Calhoun (Exp. 2) utilized a third comparison that was not related to the sample in any way that they could see. In this experiment, matching-to-sample configurations were composed of a taxonomic relation (either A–B or A–C), a taxonomic relation (A–D), and a third selection ostensibly not related to the sample (A–/–E). On these trial types the young children of that experiment, rarely selected the unrelated comparison. This outcome suggests that the third comparison functionally was in no relation to the sample, that the other comparisons were, and that the sample instructed comparison selection that evidenced one of the two

known relations. Therefore, although the present experiments contained no conditions that prove conditionality, generalization from similar experiments which used similar subjects (college students) and similar tasks supports the inference that the task herein is functionally a conditional discrimination.

REFERENCES

- Baldwin, D. (1992). Clarifying the role of shape in children's taxonomic assumption. *Journal of Experimental Child Psychology*, 54, 392–416.
- Bauer, P. J., & Mandler, J. M. (1989). Taxonomies and triads: Conceptual organization in one- to two-year-olds. *Cognitive Psychology*, 21, 156–184.
- Brunow, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- Bush, K. M., Sidman, M., & deRose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior*, 51, 29–45.
- Carrigan, P., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior*, 58, 183–204.
- Carter, D. E., & Werner, T. J. (1978). Complex learning and information processing by pigeons: A critical analysis. *Journal of the Experimental Analysis of Behavior*, 29, 565–601.
- Cumming, W. W., & Berryman, R. (1965). The complex discriminated operant: Studies of matching to sample and related problems. In D. J. Mostofsky (Ed.), *Stimulus generalization* (pp. 284–330). Stanford, CA: Stanford University Press.
- Denney, N. (1974). Evidence for developmental changes in categorization criteria for children and adults. *Human Development*, 17, 41–53.
- Denney, N., & Ziobrowski, M. (1972). Developmental changes in clustering criteria. *Journal of Experimental Child Psychology*, 13, 275–282.
- Dube, W. V., & McIlvane, W. J. (1996). Some implications of a stimulus control topography analysis for emergent behavior and stimulus classes. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in*

- humans and animals* (pp. 197–218). Amsterdam: Elsevier.
- Dunham, P., & Dunham, F. (1995). Developmental antecedents of taxonomic and thematic strategies at 3 years of age. *Developmental Psychology*, 31, 483–493.
- Fenson, L., Cameron, M. S., & Kennedy, M. (1988). Role of perceptual and conceptual similarity in category matching at age two years. *Child Development*, 59, 897–907.
- Gatch, M. B., & Osborne, J. G. (1989). Transfer of contextual stimulus function via equivalence class development. *Journal of the Experimental Analysis of Behavior*, 51, 369–378.
- Gelman, R. (1978). Cognitive development. *Annual Review of Psychology*, 29, 297–332.
- Gelman, R., & Greeno, J. G. (1989). On the nature of competence: Principles for understanding in a domain. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 125–186). Hillsdale, NJ: Erlbaum.
- Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 50, 461–467.
- Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child*. New York: Norton.
- Lowenkron, B. (1998). Some logical functions of joint control. *Journal of the Experimental Analysis of Behavior*, 69, 327–354.
- Markman, E. M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 72–106). NY: Cambridge University Press.
- Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, 16, 1–27.
- McIlvane, W. J., & Dube, W. V. (1992). Stimulus control shaping and stimulus control topographies. *The Behavior Analyst*, 89–94.
- Osborne, J. G., & Calhoun, D. O. (1998). Themes, taxons, and trial types in children's matching to sample: Methodological considerations. *Journal of Experimental Child Psychology*, 68, 35–50.
- Osborne, J. G., & Koppel, L. (2001). Acquisition, generalization, and contextual control of taxonomic and thematic relational responding. *The Psychological Record*, 51, 185–205.
- Premack, D. (1983). The codes of man and beasts. *Behavioral and Brain Sciences*, 6, 125–167.
- Ray, B. (1969). Selective attention: the effects of combining stimuli which control incompatible behavior. *Journal of the Experimental Analysis of Behavior*, 12, 539–550.
- Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Advances and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.
- Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, 74, 127–146.
- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37, 5–22.
- Smiley, S. S., & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology*, 28, 249–257.
- Stromer, R., & Osborne, J. G. (1982). Control of adolescents' arbitrary matching-to-sample by positive and negative stimulus relations. *Journal of the Experimental Analysis of Behavior*, 37, 329–348.
- Vygotsky, L. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Waxman, S. R., & Namy, L. L. (1997). Challenging the notion of a thematic preference in young children. *Developmental Psychology*, 33, 555–567.