JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2002, **78,** 291–313 NUMBER 3 (NOVEMBER)

THE FORMATION OF A GENERALIZED CATEGORIZATION REPERTOIRE: EFFECT OF TRAINING WITH MULTIPLE DOMAINS, SAMPLES, AND COMPARISONS

LANNY FIELDS, KENNETH F. REEVE, PRIYA MATNEJA, ANTONIOS VARELAS,
JAMES BELANICH, ADRIENNE FITZER, AND KIM SHAMOUN

QUEENS COLLEGE AND
THE GRADUATE CENTER OF CUNY AND
CALDWELL COLLEGE

The present experiment explored the effects of three variables on the spontaneous categorization of stimuli in perceptually distinct and novel domains. Each of six stimulus domains was created by morphing two images that were the domain endpoints. The endpoints of the domains were male and female faces, two abstract drawings, a car and a truck, two banded-elevation satellite land images, a tree and a cat, and two false-color satellite images. The stimulus variants at each end of a domain defined two potential perceptual classes. Training was conducted in a matching-to-sample format and used stimuli from one or two domains, one or three variants per class as samples, and one or three variants per class as comparisons. The spontaneous categorization of stimuli in the untrained stimulus domains showed the emergence of a generalized categorization repertoire. The proportion of spontaneously categorized stimuli in the new domains was positively related to the number of domains and samples used in training, and was inversely related to the number of comparisons used in training. Differential reaction times demonstrated the discriminability of the stimuli in the emergent classes. This study is among the first to provide an empirical basis for a behavior-analytic model of the development of generalized categorization repertoires in natural settings.

Key words: perceptual classes, spontaneous categorization, matching to sample, selection-based responding, college students

The phenomenon of categorization has a rich and varied history in psychology and has been studied under the rubrics of concept formation or stimulus class formation. One type of category, called a *perceptual class,* contains stimuli that can be arrayed along some physically, mathematically, or psychometrically defined dimension (Bourne, Dominowski, & Loftus, 1979; Fields & Reeve, 2000; Herrnstein, 1990; Hrycenko & Harwood, 1980; Lea & Ryan, 1984). At its simplest level, a perceptual class or category is said to emerge when a group of stimuli satisfy three criteria. First, all of the stimuli in the set must occasion the same response after it is trained to occur in the presence of only some of the stimuli in the set. Second, that response should occur with very low probabilities in the presence of stimuli that are in a different region of the same dimension or in other stimulus domains. Third, many of the stimuli in the set must be discriminable from each other (Cook, Wright, & Kendrick, 1990; Fields & Reeve, 2001; Keller & Schoenfeld, 1950; Lea, 1984; Reeve & Fields, 2001; Wasserman, Kiedinger, & Bhatt, 1988). Behavior indicative of control by a perceptual class, then, enables an individual to respond appropriately to the inevitable variations that occur among the exemplars of a class in natural settings. Thus, the formation of perceptual classes is of substantial adaptive utility (Bruner, Goodnow, & Austin, 1965; Medin & Smith, 1984; Rosch & Mervis, 1975; Smith, 1989).

First explored experimentally by Hull (1920), perceptual classes have been formed most frequently using multiple-exemplar training. This involves the presentation of many exemplars and nonexemplars along with differential reinforcement of responding to the stimuli in the two sets. A number of studies have shown that the likelihood of forming perceptual classes is a positive func-

---

This research was conducted with support from Contract DASW01-96-K-0009 from the U.S. Army Research Institute, and by PSC-CUNY Research Awards 68547, 69567, and 61617. We thank Mari Watanabe for her critical editorial comments during the preparation of the final manuscript and Xiqiang Zhu for his assistance in the development of the software used to conduct the experiment and analyze the data reported herein.

Reprints can be obtained from Lanny Fields, Department of Psychology, Queens College/CUNY, 65-30 Kissena Boulevard, Flushing, New York 11367. Upon request, the authors will forward a PDF file that contains the individual generalization gradients for all tests, domains, and subjects.

tion of the number of within-class exemplars used during discrimination training. Examples of this approach can be found in the behavior-analytic literature (Bhatt & Wright, 1992; Cook et al., 1990; Malott & Siddall, 1972; Pluchino, 1997; Wright, Cook, Rivera, Sands, & Delius, 1988) and the cognitive literature (Homa & Chambliss, 1975; Homa, Cross, Cornell, Goldman, & Swartz, 1973; Homa & Little, 1985; Homa, Sterling, & Treple, 1981; Omohundro, 1981). In general, multiple-exemplar training induces control of behavior by a set of perceptually similar stimuli through two processes: (a) the establishment of a predictive relation between the availability of reinforcement and the stimuli in a given region of a domain, and (b) the elimination of a predictive relation between availability of reinforcement and the idiosyncratic features of the stimuli used as positive and negative exemplars (Fields & Reeve, 2001; Keller & Schoenfeld, 1950; Tiemann & Markle, 1990; Wright et al., 1988).

These multiple-exemplar training procedures have been used to establish perceptual classes in typically functioning adults and children (Brown, Brown, & Poulson, 1995; Herrnstein & de Villiers, 1980; Homa & Chambliss, 1975; Homa & Little, 1985; Njegovan, Ito, Mewhort, & Weisman, 1995; Wasserman & DeVolder, 1993), people with mental retardation or autism (Gena, Krantz, McClannahan, & Poulson, 1996; Goldstein & Mousetis, 1989; Handleman, 1979; McIlvane, Dube, Green, & Serna, 1993; Stokes, Baer, & Jackson, 1974; Young, Krantz, McClannahan, & Poulson, 1994), primates (Bhatt & Wright, 1992; Cook et al., 1990; Wright et al., 1988), and pigeons (Cerella, 1979; Cook et al., 1990; Herrnstein, Loveland, & Cable, 1976; Honig & Stewart, 1988; Porter & Neuringer, 1985; Wright et al., 1988). Experimentally generated classes have employed many perceptual domains that include line length (Fields, Reeve, Adams, Brown, & Verhave, 1997); tonal range (Cross & Lane, 1962; Njegovan et al., 1995); fill patterns (Reeve & Fields, 2001); trees and water (Herrnstein et al., 1976; Malott & Siddall, 1972); leaves (Cerella, 1979); cars, cats, and flowers (Wasserman et al., 1988); fish (Herrnstein & de Villiers, 1980); people (Malott & Siddall, 1972); locations on a college campus (Honig & Stewart, 1988); manufactured objects (Lubow, 1974); musical styles (Porter & Neuringer, 1995); movement (Dittrich & Lea, 1993); and computer-generated animal pictures (Blough, 1990; Jitsumori, 1996).

Casual observation in natural settings suggests that many individuals not only categorize stimuli in domains with which they have had ''training,'' but also spontaneously categorize stimuli in other domains without such training. Such a repertoire would exemplify *generalized categorization.* Because a generalized categorization repertoire transcends the definitional particulars of stimuli in any given domain, it would be of greater adaptive significance than the ability to form a particular stimulus class alone. Indeed, the absence of such a generalized categorization repertoire may characterize some of the behavioral deficits shown by individuals with autism, mental retardation, or dementia (Dube, Iennaco, Rocco, Kledaras, & McIlvane, 1992). The fact that an individual categorizes stimuli in a given domain, however, does not necessarily imply that the individual *will* spontaneously categorize stimuli in new domains. It is difficult to imagine the emergence of a generalized categorization repertoire in any type of individual without some prior history of reinforcement involving exemplars and nonexemplars. Although it seems likely that such repertoires are established in natural settings by specific reinforcement histories, we are not aware of experiments that have studied the formation of such repertoires. The purpose of the present experiment therefore was to examine a number of variables that may influence the formation of a generalized categorization repertoire.

In many studies, multiple-exemplar training has been used to induce generalized repertoires that do not qualify as generalized categorization. Relevant topics of investigation include learning set (Harlow, 1949), identity matching (Brown et al., 1995; Wright et al., 1988), mirror-image symmetry (Delius & Haber, 1978), associative symmetry (Boelens & Van Den Broek, 2000), and imitation (Young et al., 1994). Multiple-exemplar training also has been used to establish generalized discriminative control over relational frames, such as *greater than, less than, opposite,* and *different,* that humans presumably acquire preexperimentally (Hayes & Barnes, 1997; O'Hora, Roche, Barnes-Holmes, & Smeets, 2002;

Steele & Hayes, 1991). By implication, then, a generalized categorization repertoire might emerge after some form of multiple-exemplar training.

As mentioned above, the formation of perceptual classes has been most commonly accomplished by using multiple-exemplar training and testing. When used in the context of matching-to-sample procedures, the many exemplars, to our knowledge, always have been presented as samples along with a single comparison per class. Thus, one would predict that the categorization of stimuli in new domains might also be directly related to the number of different stimuli used as samples in class formation training.

The goal of class formation training is to establish relations among all exemplars in a set of perceptually similar stimuli. Logically, this ought to be maximized by the training of relations among the variants in a set on a bidirectional basis. This would involve the presentation of many exemplars in a set both as samples and as comparisons. Thus, all of the variants used in training are likely to become related to each other regardless of their behavioral functions during training. Thus, induction of a generalized categorization repertoire should be maximized by such a training history. We have not found any studies, however, that explored the effects of such a training procedure on the formation of perceptual classes or the induction of a generalized categorization repertoire.

When training is conducted with stimuli drawn from only one domain, class-indicative responding is likely to come under the contextual control of the stimuli in that domain alone, because those are the only stimuli correlated with reinforcement. Therefore, little spontaneous categorization of stimuli in new domains would be expected after single-domain training. When training is conducted with stimuli in perceptually different domains, however, the idiosyncratic features of one domain are less likely to become predictive of reinforcement. Therefore, it is much less likely that the categorization of stimuli will come under the contextual control of stimuli in a single domain. Rather, it is more likely that categorization-indicative responding will generalize to stimuli in new domains (Fields & Reeve, 2001). Thus, the categorization of stimuli in new domains should be directly related to the number of different domains used for class formation training.

The purpose of the present study, then, was to investigate how multiple-exemplar manipulations applied to three components of training influenced the emergence of a generalized categorization repertoire. Subjects were presented with stimuli from six perceptually distinct domains. Training was conducted using a matching-to-sample procedure with (a) one or three class members as samples, (b) one or three class members as comparisons, and (c) stimuli in one or two domains. Training was followed by a generalization test conducted with the stimuli from the remaining domains. The performances in the generalization test showed how the parameters of training influenced the categorization of novel stimuli in the test domains and, thus, the formation of a generalized categorization repertoire.

METHOD

Subjects

Thirty-six undergraduate students at Queens College/CUNY participated. They were recruited from advanced psychology classes and reported no familiarity with the research area. These subjects received partial course credit upon completion of the experiment, but credit did not depend upon performance. All subjects were required to read and sign an informed consent statement prior to participation. The entire experiment lasted for about 3 hr per subject and was completed in one or two sessions. The 9 subjects in each of the four experimental groups were assigned randomly.

Apparatus and Stimuli

The experiment was conducted with personal computers that displayed all stimuli on 15-in. SVGA color monitors. Images were projected on an 800 × 600 pixel array in which each pixel had a 0.25-in. dot pitch. Responses consisted of touching specific keys on a standard keyboard. The experiment was controlled by customized software.

Stimuli in six domains (referred to by the letters A through F) were used in the experiment. Domains A through F were called female–male, abstract pictures, truck–car, N. Korea–Germany, tree–cat, and Bosnia–Cuba, respectively. The images illustrated in Figure 1 depict the endpoints of each domain. The actual satellite images were projected in RGB 24-bit color on the computer monitors.

Stimuli that varied systematically between the endpoint stimuli in a domain were generated using commercially available morphing software (Figuracion, 1998). The software superimposes the endpoints of the domain on each other and systematically changes their relative salience, thereby producing a continuum of 500 variants between the two endpoints. Figure 2 illustrates the results of this process with some representative variants from the female–male domain.

Procedure

Definition of morph-based perceptual classes. In the progression of morph-based stimuli, many of the variants at one end of a domain resemble each other. Likewise, many of the variants at the other end of a domain resemble each other. The variants at the opposite ends of a domain, however, do not resemble each other. Thus, the stimuli at each end of a domain may be said to compromise a potential perceptual class. The endpoints of a domain serve as the *anchor* stimuli (a) of the potential classes at each end of the domain, and were designated X1a and X2a, respectively (where X denotes the domain and the numbers refer to the classes). The variant most distant from the anchor that was viewed as being perceptually related to the anchor was referred to as the *boundary* stimulus (b) of the class. The boundaries of Classes 1 and 2 were denoted X1b and X2b, respectively. The anchor and boundary stimuli at one end of a domain, then, defined the range of variants that constituted a perceptual class. In addition, we identified the variant that appeared to be perceptually equidistant from the anchor and boundary of Classes 1 and 2. These variants were referred to as the *midpoints* (m) of the classes, and were designated as X1m and X2m, respectively. Finally, we identified the variant that was perceived to be perceptually equidistant from the boundaries of the two classes at the ends of a domain. That variant, referred to as the *neither* stimulus (n) for that domain, was designated as Xn, and did not serve as a member of either Class 1 or Class 2. Figure 3 contains the anchor, midpoint, and boundary stimuli for potential Classes 1 and 2 and the neither stimulus for the domain. The procedures used to identify these five variants in each of the six domains used in the experiment are described in the Appendix.

Trial format and responses within a trial. The experiment was conducted in a trial-by-trial matching-to-sample format. The sample and comparison stimuli were presented on the computer screen in 2.5-cm squares arranged in an equilateral triangular array, with the sample at the apex and the comparisons at the corners of the base of the array. A trial began when ''press ENTER'' appeared on the screen. Pressing the enter key cleared the screen and displayed a sample stimulus at the top center of the monitor. Pressing the space bar displayed two comparison stimuli at the bottom left and right corners of the monitor while the sample remained on the computer monitor. Phase 1 of the experiment contained these two comparisons only. All remaining phases of the experiment also contained the neither option. This was represented by the phrase ''If NEITHER press 4,'' which appeared on the computer monitor between the two other comparisons.

During a trial, the left or right comparison was selected by pressing the 1 or 2 key, respectively. The neither comparison was selected by pressing the 4 key. A comparison selection cleared the screen and displayed a feedback message centered on the screen. When informative feedback was scheduled during training trials, the messages ''RIGHT'' or ''WRONG'' appeared, depending on the accuracy of the comparison selection. The message remained on the screen until the R (for right) or W (for wrong) key was pressed to terminate the trial. On some training and all test trials, trial termination was signaled by the presentation of a message that provided noninformative feedback regarding comparison selection. The message consisted of ''- -E- -,'' which remained on the computer screen until the subject pressed the E key. This was an observing response to the noninformative feedback message. After an appropriate observing response was made, the screen was cleared and the next trial began (Fields, Landon-Jimenez, Buffington, & Adams, 1995).

Trial block structure and feedback contingencies. Each phase of training and testing was con-

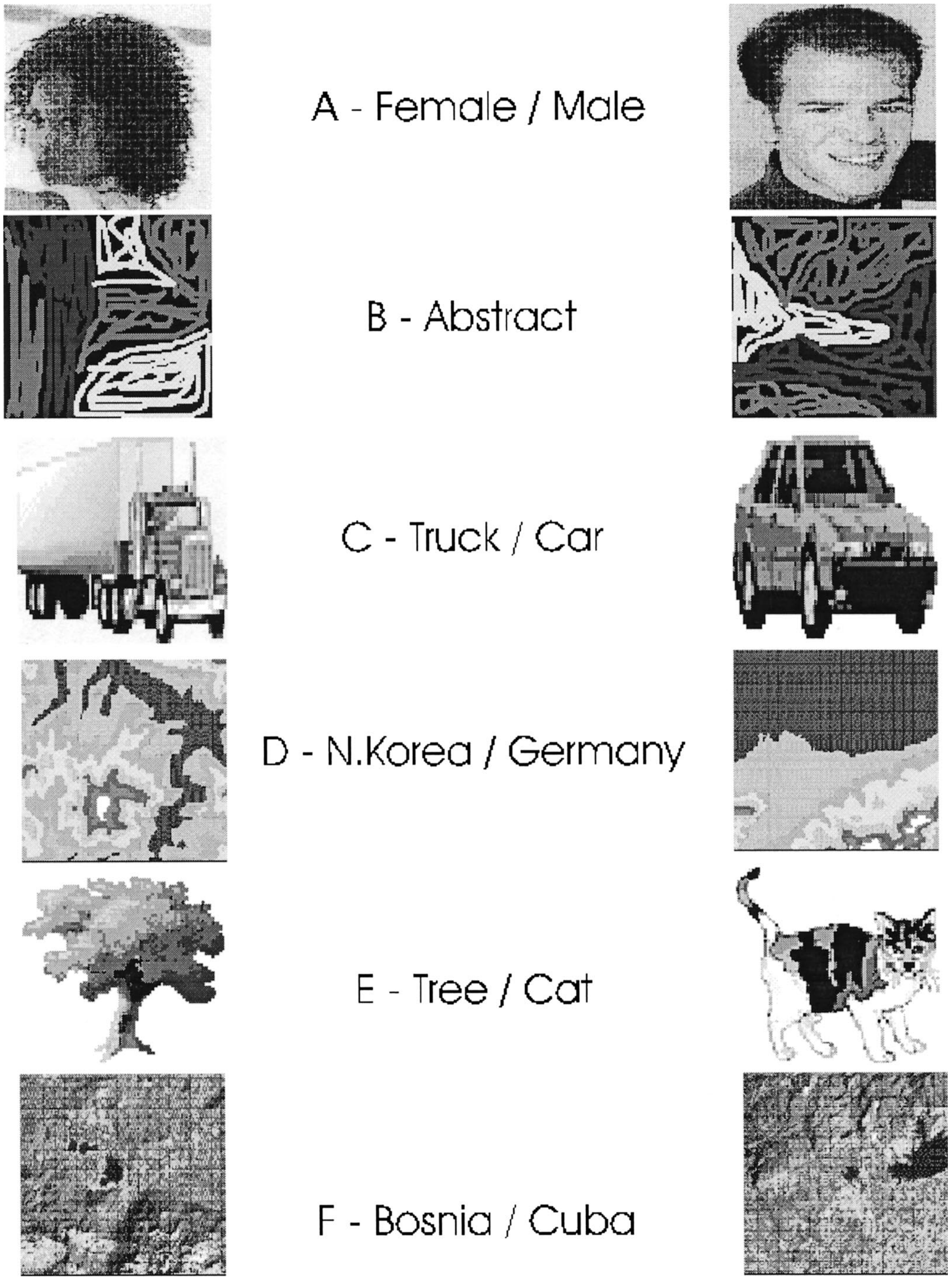


Fig. 1. The anchor stimuli that defined the endpoints of the six domains used in the experiment. Each domain is identified by a name (e.g., female–male) and a letter (e.g., A).

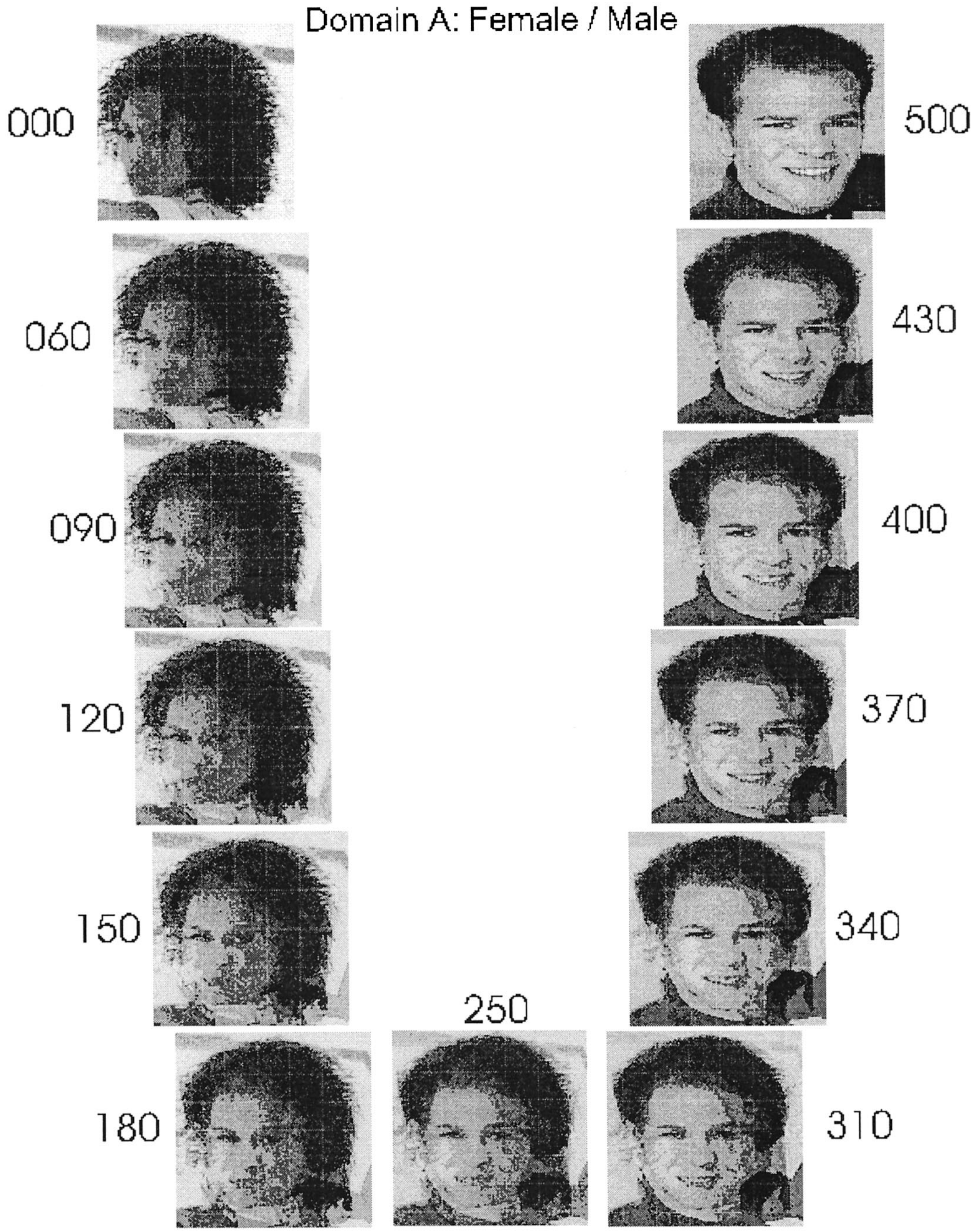


Fig. 2. A series of morphed images that represent intermediate stimuli between the endpoints of the female–male morphed stimulus domain. The anchor stimuli are at the top left and top right. The numerals next to the stimuli are the values assigned to them by the morphing program.

Fig. 3. All stimuli used in the experiment. Each column contains the stimuli in a given domain indicated at the top of the column. The anchor, midpoint, and boundary stimuli for Classes 1 and 2 are illustrated in the top three and bottom three rows, respectively. The neither stimuli for each domain are illustrated in the middle row. The anchor, midpoint, and boundary stimuli for Class 1 in a domain are represented as X1a, X1m, and X1b, respectively. The anchor, midpoint, and boundary stimuli for Class 2 in a domain are represented as X2a, X2m, and X2b, respectively. The neither stimulus in a domain is represented as Xn. The symbolic representation for each stimulus is completed by substitution of the domain letter for X.

ducted in blocks of trials. The trials in a block were presented in a random order without replacement. At the start of training, a block was presented repeatedly with informative feedback following each comparison selection until all trials within the block occasioned 100% correct responding. Thereafter, the percentage of trials that occasioned informative feedback was reduced to 75%, 25%, and finally 0% over successive blocks as long as 100% accuracy within a block was maintained. During feedback reduction, the trials that were followed by informative feedback were randomly determined. If 100% correct responding was not achieved within three blocks at a given feedback level during training, the subject was returned to the previous feedback level.

Phase 1: Instructions and keyboard familiarization. Prior to the experiment, subjects were presented with the following instructions on the screen:

> Thank you for volunteering to participate. PLEASE DO NOT TOUCH ANY KEYS ON THE KEYBOARD YET! You will be presented with many trials. Each trial contains three or four CUES that are shapes, symbols, or common words. YOUR TASK IS TO DISCOVER HOW TO RESPOND CORRECTLY TO THE CUES BY PRESSING CERTAIN KEYS ON THE COMPUTER'S KEYBOARD. Initially, INSTRUCTIONS will tell you how to respond to the cues, and LABELS will help you identify the cues on the screen. The labels and instructions will slowly disappear. The experiment is conducted in phases. When each phase ends, the computer will sometimes tell you how you did. If you want to take a break at any time, call the experimenter. Thank you for your cooperation! Press the space bar to continue.

After pressing the space bar, subjects learned to emit the appropriate keyboard responses to complete a trial. To accomplish this, 16 trials, each containing three English words such as *king, queen,* and *camel,* were presented. The semantic relatedness between the sample word (e.g., *king*) and one of the comparisons (e.g., *queen*) was used to prompt the selection of the correct comparison. Informative feedback followed each comparison selection (see Fields et al., 1997, for further details).

Correct responding to the stimuli in a trial during Phase 1 also was facilitated by instructional prompts (e.g., ''Make your choice by pressing 1 or 2'') that were deleted in a serial manner across trials (see Fields et al., 1997, or Fields, Adams, Verhave, & Newman, 1990, for further details). Phase 1 ended once the stimuli were presented without prompts and performance exceeded 87% accuracy (14 of 16 correct trials) during a single block. In the remaining phases, the instruction used to prompt the appropriate key press during Phase 1 reappeared on the screen for three subsequent trials whenever a subject pressed a nonexperimentally defined key during a trial.

Phase 2: Training. All training was conducted in a matching-to-sample format. The samples were drawn from the two potential classes in a domain. In some conditions, the samples were the anchor stimuli alone. In other conditions, the samples consisted of the anchor, midpoint, and boundary stimuli. These stimuli were presented in a randomized order across trials. The selection of the comparison drawn from the same potential class as the sample occasioned the ''RIGHT'' message. The selection of the other comparisons occasioned the presentation of the ''WRONG'' feedback message. Depending on the experimental condition, the comparisons in a trial consisted of the pair of anchor stimuli from the two potential classes in one domain, or pairs of anchor, midpoint, and boundary stimuli from the two potential classes in a domain. All training and testing conditions also included the presentation of the neither stimulus as a sample. Selection of the neither comparison (pressing the 4 key) in the presence of the neither stimulus for that domain occasioned the presentation of the ''RIGHT'' feedback message. Selection of either of the other comparisons occasioned the presentation of the ''WRONG'' feedback message.

Four groups of subjects were studied to determine how the spontaneous categorization of stimuli in new domains was influenced by the number of domains, the number and type of stimuli used as samples, and the number and type of stimuli used as comparisons in training. The following notational system specifies each training condition. Experimental condition is denoted by a series of one or more capital letters followed by a parenthesis that includes a string of lowercase letters separated by a dash. The number of domains used for training are indicated by the number of uppercase letters (A, B, or both) and the letters designate the particular domains used in training. The lowercase letters designate the sample and comparison stimuli with which training was conducted in each of the domains used for training (a = anchor, m = midpoint, and b = boundary). The letters that precede and follow the dash represent the stimuli used as samples and comparison pairs in training, respectively. Using this notation, the four conditions were designated as A(a-a) training, A(amb-a) training, AB(amb-a) training, and AB(amb-amb) training. The details of each training condition are described below. A comparison of the outcomes of these conditions shows how the emergence of a generalized categorization repertoire is influenced by (a) the number of domains used in training, (b) the number of sample stimuli used in training, and (c) the number of comparison stimuli used in training.

A(a-a) training. In the A(a-a) condition, only stimuli in the A domain were used for training. The anchor stimuli from both ends of the A domain were presented as comparisons on all trials along with the neither comparison. The samples used for training were the anchor stimuli at each end of the A domain along with the neither stimulus in that domain. Differential feedback was presented for the selection of a comparison that was identical to the sample, and for the selection of the neither comparison in the presence of the neither stimulus presented as the sample. No training was conducted with the midpoint and boundary stimuli in each potential class drawn from the A domain. Thus, in the A(a-a) condition, identity conditional discriminations were established with the anchor stimuli using single-exemplar training with the sample stimuli from one domain only. The A(a-a) condition can also be called single-domain one-to-one training. The number of trials and stimulus configurations in the training trials are indicated in Table 1.

A(amb-a) training. In the A(amb-a) condition, only stimuli in the A domain were used for training. The anchor, midpoint, and boundary stimuli from both ends of the domain were presented as samples. The comparisons on all trials were the anchor stimuli at the ends of the A domain and the neither comparison. The anchor, midpoint, and boundary stimuli from one end of the domain occasioned differential feedback for the selection of the anchor stimulus from the end of the same domain. These contingencies were used with stimuli at both ends of each domain. In addition, the presentation of the neither stimulus from the A domain occasioned differential feedback for the selection of the neither option. Thus, multiple-exemplar training was conducted with the sample stimuli from one domain only. It was not conducted with the remaining domains or with other comparison stimuli. A(amb-a) training can also be called single-domain many-to-one training. The number of trials and stimulus configurations in the training trials are indicated in Table 1.

AB(amb-a) training. In the AB(amb-a) condition, training was conducted with stimuli from the A domain first, and then with stimuli from the B domain. As in the previously described A(amb-a) condition, the anchor, midpoint, and boundary stimuli from one end of the domain occasioned the presentation of differential feedback for the selection of the anchor stimulus from the same end of the domain. These contingencies were used with stimuli at both ends of each domain. In addition, the presentation of the neither stimuli from the A and B domains occasioned differential feedback for the selection of the neither option. Thus, multiple-exemplar training was conducted with the sample stimuli from two domains. It was not conducted with different comparison pairs in a domain. AB(amb-a) training can also be called double-domain many-to-one training. The number of trials and stimulus configurations in the training trials are indicated in Table 1.

For each domain (A and B), training was conducted in three stages. In Stage 1, the anchor and neither stimuli were used as samples. In Stage 2, the anchor, midpoint, and neither stimuli were presented as samples. In Stage 3, the anchor, midpoint, boundary, and neither stimuli were presented as samples. Stages 1 through 3 were conducted with the presentation of differential feedback on all trials. Once completed, the trials in Stage 3 were presented under conditions of reduced feedback.

AB(amb-amb) training. Finally, in the AB(amb-amb) condition, training was again conducted serially with stimuli from the A

Table 1

A symbolic representation of the stimulus configurations used in each training condition and in testing. Sa = sample, Co = comparison, NC = neither comparison. The letters a, m, and b symbolize the anchor, midpoint, and boundary stimuli, respectively. The numbers 1 and 2 represent Classes 1 and 2, respectively. The capital letters that precede 1 and 2 symbolize the domains from which the stimuli are used. The value in the Trials column indicates the number of times each sample–comparison configuration was presented in a training or testing block. In each block, each comparison was presented equally often on the left and right. The X in the generalization test section represents Domains A through F.

Condition	Sa	Co+	Co−	Co−	Trials	Sa	Co+	Co−	Co−	Trials
A(a-a)	A1a	A1a	NC	A2a	4	A2a	A2a	NC	A1a	4
	An	NC	A1a	A2a	6					
A(amb-a)	A1a	A1a	NC	A2a	4	A2a	A2a	NC	A1a	4
	A1m	A1a	NC	A2a	4	A2m	A2a	NC	A1a	4
	A1b	A1a	NC	A2a	4	A2b	A2a	NC	A1a	4
	An	NC	A1a	A2a	4					
AB(amb-a)										
Stage 1	A1a	A1a	NC	A2a	4	A2a	A2a	NC	A1a	4
	An	NC	A1a	A2a	4					
	B1a	B1a	NC	B2a	4	B2a	B2a	NC	B1a	4
	Bn	NC	B1a	B2a	4					
Stage 2	A1a	A1a	NC	A2a	4	A2a	A2a	NC	A1a	4
	A1m	A1a	NC	A2a	4	A2m	A2a	NC	A1a	4
	An	NC	A1a	A2a	4					
	B1a	B1a	NC	B2a	4	B2a	B2a	NC	B1a	4
	B1m	B1a	NC	B2a	4	B2m	B2a	NC	B1a	4
	Bn	NC	B1a	B2a	4					
Stage 3	A1a	A1a	NC	A2a	4	A2a	A2a	NC	A1a	4
	A1m	A1a	NC	A2a	4	A2m	A2a	NC	A1a	4
	A1b	A1a	NC	A2a	4	A2b	A2a	NC	A1a	4
	An	NC	A1a	A2a	4					
	B1a	B1a	NC	B2a	4	B2a	B2a	NC	B1a	4
	B1m	B1a	NC	B2a	4	B2m	B2a	NC	B1a	4
	B1b	B1a	NC	B2a	4	B2b	B2a	NC	B1a	4
	Bn	NC	B1a	B2a	4					
AB(amb-amb)										
Stage 1	A1a	A1a	NC	A2a	2	A2a	A2a	NC	A1a	2
	A1m	A1a	NC	A2a	2	A2m	A2a	NC	A1a	2
	A1b	A1a	NC	A2a	2	A2b	A2a	NC	A1a	2
	An	NC	A1a	A2a	6					
	B1a	B1a	NC	B2a	2	B2a	B2a	NC	B1a	2
	B1m	B1a	NC	B2a	2	B2m	B2a	NC	B1a	2
	B1b	B1a	NC	B2a	2	B2b	B2a	NC	B1a	2
	Bn	NC	B1a	B2a	6					
Stage 2	A1a	A1m	NC	A2m	2	A2a	A2m	NC	A1m	2
	A1m	A1m	NC	A2m	2	A2m	A2m	NC	A1m	2
	A1b	A1m	NC	A2m	2	A2b	A2m	NC	A1m	2
	A1a	A1b	NC	A2b	2	A2a	A2b	NC	A1b	2
	A1m	A1b	NC	A2b	2	A2m	A2b	NC	A1b	2
	A1b	A1b	NC	A2b	2	A2b	A2b	NC	A1b	2
	An	NC	A1m	A2m	6					
	An	NC	A1b	A2b	6					
	B1a	B1m	NC	B2m	2	B2a	B2m	NC	B1m	2
	B1m	B1m	NC	B2m	2	B2m	B2m	NC	B1m	2
	B1b	B1m	NC	B2m	2	B2b	B2m	NC	B1m	2
	B1a	B1b	NC	B2b	2	B2a	B2b	NC	B1b	2
	B1m	B1b	NC	B2b	2	B2m	B2b	NC	B1b	2
	B1b	B1b	NC	B2b	2	B2b	B2b	NC	B1b	2

Table 1

(*Continued*)

Condition	Sa	Co+	Co−	Co−	Trials	Sa	Co+	Co−	Co−	Trials
	Bn	NC	B1m	B2m	6					
	Bn	NC	B1b	B2b	6					
Stage 3	A1a	A1a	NC	A2a	2	A2a	A2a	NC	A1a	2
	A1m	A1a	NC	A2a	2	A2m	A2a	NC	A1a	2
	A1b	A1a	NC	A2a	2	A2b	A2a	NC	A1a	2
	A1a	A1m	NC	A2m	2	A2a	A2m	NC	A1m	2
	A1m	A1m	NC	A2m	2	A2m	A2m	NC	A1m	2
	A1b	A1m	NC	A2m	2	A2b	A2m	NC	A1m	2
	A1a	A1b	NC	A2b	2	A2a	A2b	NC	A1b	2
	A1m	A1b	NC	A2b	2	A2m	A2b	NC	A1b	2
	A1b	A1b	NC	A2b	2	A2b	A2b	NC	A1b	2
	An	NC	A1a	A2a	6					
	An	NC	A1m	A2m	6					
	An	NC	A1b	A2b	6					
	B1a	B1a	NC	B2a	2	B2a	B2a	NC	B1a	2
	B1m	B1a	NC	B2a	2	B2m	B2a	NC	B1a	2
	B1b	B1a	NC	B2a	2	B2b	B2a	NC	B1a	2
	B1a	B1m	NC	B2m	2	B2a	B2m	NC	B1m	2
	B1m	B1m	NC	B2m	2	B2m	B2m	NC	B1m	2
	B1b	B1m	NC	B2m	2	B2b	B2m	NC	B1m	2
	B1a	B1b	NC	B2b	2	B2a	B2b	NC	B1b	2
	B1m	B1b	NC	B2b	2	B2m	B2b	NC	B1b	2
	B1b	B1b	NC	B2b	2	B2b	B2b	NC	B1b	2
	Bn	NC	B1a	B2a	6					
	Bn	NC	B1m	B2m	6					
	Bn	NC	B1b	B2b	6					
Testing X(amb-a)	X1a	X1a	NC	X2a	2,3	X2a	X2a	NC	X1a	2
	X1m	X1a	NC	X2a	2	X2m	X2a	NC	X1a	2
	X1b	X1a	NC	X2a	2	X2b	X2a	NC	X1a	2
	Xn	NC	X1a	X2a	2					

and B domains. For a given domain, the anchor, midpoint, and boundary stimuli of both potential classes and the neither stimuli from these domains were presented as samples. In addition, matched pairs of comparison stimuli were presented: Anchors 1 and 2 (X1a and X2a), Midpoints 1 and 2 (X1m and X2m), or Boundaries 1 and 2 (X1b and X2b). Differential feedback was presented for the selection of a comparison from the same potential class as the sample. In addition, the presentation of the neither stimuli in the A and B domains occasioned differential feedback for the selection of the neither option. Thus, multiple-exemplar training was conducted with the sample stimuli, comparison stimuli, and stimulus domains. The AB(amb-amb) training can also be called double-domain many-to-many training. The number of trials and stimulus configurations in the training trials are indicated in Table 1.

For each domain (A and B), training was conducted in three stages. Stage 1 involved the establishment of a-a, m-a, b-a, and n-a conditional discriminations. Stage 2 involved the establishment of the a-b, m-b, b-b, a-m, m-m, b-m, n-m, and n-b conditional discriminations. Stage 3 involved the maintenance of the conditional discriminations established in the two prior stages (i.e., a-a, m-m, b-b, a-b, b-a, a-m, m-a, b-m, m-b, n-a, n-m, and n-b). Stages 1 through 3 were conducted with the presentation of informative feedback on all trials. Once completed, the trials in stage 3 were presented under conditions of reduced feedback. These stages were then repeated with stimuli from the B domain.

Phase 3: X(amb-a) generalization testing. At the completion of each training condition, a generalization test was administered to all subjects. All tests were conducted in the amb-a or many-to-one format. Each test block involved the presentation of the seven variants from all domains as sample stimuli: X1a, X1m, X1b, Xn, X2b, X2m, and X2a. Each sample stimulus was presented on two trials in the block. When the seven stimuli from one domain were presented as samples, they were always accompanied by the presentation of the anchor stimuli from the same domain as comparisons. All trials also contained the neither comparison as a third selection option. Its selection would indicate that the prevailing sample was not related to either of the prevailing comparison stimuli (Fields et al., 1997; Innis, Lane, Miller, & Critchfield, 1998; Reeve & Fields, 2001).

The test block was repeated three times, resulting in six presentations of each sample stimulus in the entire generalization test. Thus, the anchor, midpoint, and boundary stimuli from a potential class were presented 18 times in the generalization test. All trials were presented in a randomized order without replacement. The number of trials and stimulus configurations in the tests are indicated in Table 1. Performances occasioned by the stimuli in the novel domains were used to assess the induction of a generalized categorization repertoire.

Discriminability. The present experiment documented the discriminability of the stimuli in the emergent classes in the new domains by the measurement of differences in the reaction times occasioned by the anchor, midpoint, and boundary stimuli. The duration of each reaction time was measured from the onset of the comparison stimuli to the selection of one of the comparisons, regardless of comparison selection. Mean reaction times were computed separately for the anchor, midpoint, and boundary stimulus in each class that emerged during the generalization test. These means were based on the six presentations of each stimulus during the generalization test. Then, a grand mean reaction time for all of the anchor stimuli was calculated across test domains, classes, and subjects. Likewise, separate mean and grand mean reaction times were computed for the midpoint and boundary stimuli.

RESULTS

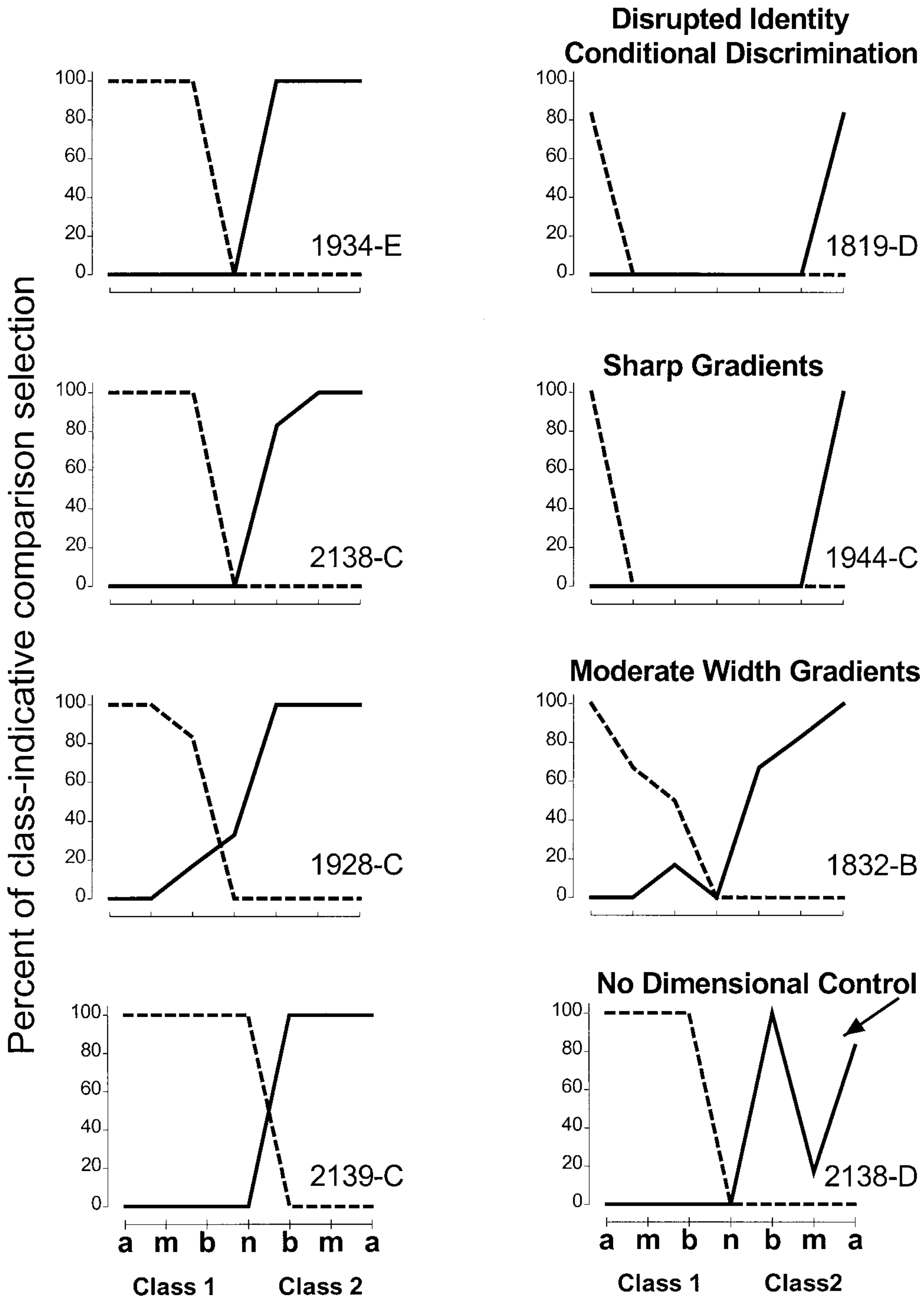
Individual Outcomes

Individual performances that indicate categorization of stimuli in new domains. This experiment determined how four training conditions and three parameters of training influenced the emergence of perceptual classes in new stimulus domains. To make that determination, however, it is first necessary to define when a set of stimuli in a test domain was functioning as a class. In the present experiment, a set of stimuli in a test domain was defined as functioning as a class when the anchor stimulus from the same end of a domain was selected on at least 17 of the 18 presentations of the anchor, midpoint, and boundary stimuli from that end of a test domain. In addition, the remaining stimuli in the domain rarely occasioned the selection of the same anchor stimulus. Other performances indicated failures to categorize stimuli in a test domain. Examples of both outcomes are illustrated in Figure 4.

The left column of Figure 4 contains some representative generalization test performances indicative of the categorization of stimuli in new domains. In the first panel, the anchor, midpoint, and boundary stimuli from each class always occasioned the selection of the related comparison, and the remaining

→

Fig. 4. Examples of primary generalization gradients obtained for individual subjects. The first three panels in the left column show generalization test performances that indicate the emergence of classes. Remaining panels show performances that are not indicative of categorization (see text for details). Each panel is identified by the subject and the domain from which the gradient was obtained. For each panel, the abscissa indicates the anchor, midpoint, and boundary stimuli for each potential class in the domain, along with the neither stimulus located between the class boundaries. The potential members of Classes 1 and 2 are located at the left and right hand sides of the abscissa. The seven stimuli on the abscissa were presented as samples during generalization tests. The ordinate indicates the relative frequency of selection of one of two anchor stimuli presented as comparisons. Dashed functions represent the relative frequency of selecting the anchor stimulus that is the same as that at the left extreme of the abscissa. Solid functions represent the relative frequency of selecting the anchor stimulus that is the same as that at the right extreme of the abscissa. If the relative frequencies of both comparison selections do not sum to 1.0, the difference represents the selection of the neither comparison.


Percent of class-indicative comparison selection
100
80
60
40
20
0
1934-E
2138-C
1928-C
2139-C
Disrupted Identity Conditional Discrimination
1819-D
Sharp Gradients
1944-C
Moderate Width Gradients
1832-B
No Dimensional Control
2138-D
a m b n b m a
Class 1
Class 2
a m b n b m a
Class 1
Class2

stimuli never occasioned the selection of the same comparison. The selection values of zero for the neither stimulus indicate that it never occasioned the selection of each endpoint of the domain. Rather, it always occasioned the selection of the neither comparison (a gradient showing the selection of the neither comparisons is not included in this and all other panels in the figure). Therefore, the two classes that emerged in the domain were functionally independent of each other (Belanich & Fields, in press; Fields et al., 1997). In the second panel, 17 of the 18 stimuli in Class 2 occasioned the selection of the related comparison, and the remaining stimuli did not. In addition, all of the stimuli in Class 1 occasioned the selection of the related comparison, and the remaining stimuli did not. In the third panel, 17 of 18 stimuli in Class 1 occasioned class-consistent comparison selections, whereas the stimuli in Class 2 always occasioned selection of the class- consistent comparison. Thus, two classes emerged. On occasion, the neither stimulus and the boundary stimulus in Class 1 also occasioned the selection of the comparison in Class 2. The neither stimulus most frequently occasioned the selection of the neither comparison. In the fourth panel, the anchor, midpoint, and boundary stimuli in Class 1, along with the neither stimulus, all occasioned the selection of the anchor stimulus in Class 1, and the remaining stimuli did not. Therefore, all four of these stimuli functioned as members of the same emergent class. In this case, the boundary stimulus of potential Class 2 occasioned the selection of the Class 1 and 2 anchor stimuli with equal likelihood, and the neither comparison was never selected. The selections of the anchor and boundary stimuli in potential Class 2 did not satisfy the criterion for class emergence.

The right column of Figure 4 contains representative generalization gradients that indicated failures to categorize the stimuli in the potential classes in the test domains. In the first panel, the anchor stimulus at each end of the domain did not always result in the selection of the corresponding anchor as a comparison. These data showed the emergence of incomplete identity conditional discriminations by both anchor stimuli. The remaining stimuli occasioned the selection of the neither comparison. In the second panel, the anchor stimuli at each end of the domain always occasioned the selection of the corresponding anchor as a comparison, whereas the remaining stimuli that were between the two anchors occasioned the selection of the neither comparison. These data showed the emergence of identity conditional discriminations and the complete discrimination of the anchor stimuli from all of the intermediate stimuli in the domain. In the third panel, the anchor at each end of the domain occasioned the selection of the corresponding anchor as a comparison, and the adjacent stimuli occasioned gradually declining likelihoods of selecting the same comparison. These moderately sloped, relatively broad generalization gradients show that the control of behavior was shared by many stimuli along the morph dimension, but control was not great enough for the stimuli to function as members of a class. In addition, the neither stimulus always occasioned the selection of the neither comparison. Finally, the midpoint stimulus at the low end of the domain infrequently occasioned the selection of the unrelated anchor stimulus from the high end of the domain. The subject always selected the neither comparison in the presence of the neither stimulus. In the fourth panel, the stimuli at the high end of the domain indicated the emergence of an incomplete identity conditional discrimination and comparison selections that were not systematically related to the value of the stimuli presented as samples. These gradients show that the dimension along which the stimuli were arrayed was not a determinant of comparison selection. In contrast, the stimuli at the low end of the domain functioned as members of an emergent class.

Number of emergent classes in new domains. Figure 5 shows the number of classes that emerged in the test domains for each subject after exposure to each training condition. Each data point represents the performance obtained for a separate subject. After A(a-a) or single-domain one-to-one training, very few subjects showed the emergence of any classes in the new domains. There was little intersubject variation in performance. After A(amb-a) or single-domain many-to-one training, some subjects showed the emergence of some classes in the new domains. In addition, there was substantial intersubject

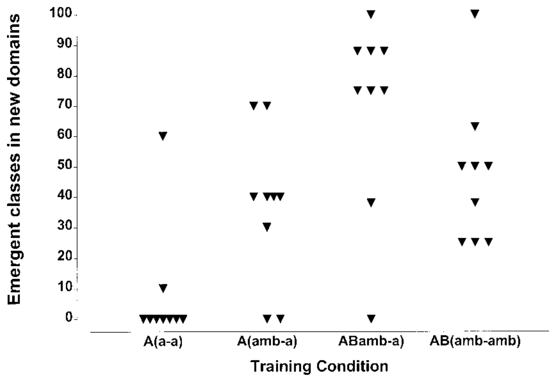


Fig. 5. The percentage of classes that emerged for the stimuli in the test domains as a function of training condition. Each data point represents an individual subject.

variation in the test performances. After AB(amb-a) or double-domain many-to-one training, many subjects showed the emergence of many classes in the new domains. In addition, there was relatively little intersubject variation in the test performances. After AB(amb-amb) or double-domain many-to-many training, some subjects showed the emergence of some classes in the new domains. In addition, there was relatively little intersubject variation in the test performances. An analysis of variance (ANOVA) showed that there was a significant difference in the effects of the four training conditions, $F(3, 32) = 29.4$, $p < .0001$.

Group Outcomes

Prevalence of emergent classes across subjects. Comparisons of the outcomes of the four training conditions show the potential effects of three training variables on the formation of a generalized categorization repertoire. A comparison of the effects of A(a-a) and A(amb-a) training showed how single- and multiple-exemplar training with sample stimuli influenced the emergence of a generalized categorization repertoire. A comparison of the effects of A(amb-a) and AB(amb-a) training showed how single- and multiple-domain training influenced the emergence of a generalized categorization repertoire. Finally, a comparison of the effects of AB(amb-a) and AB(amb-amb) training showed how single- and multiple-comparison training influenced the emergence of a generalized categorization repertoire.

Figure 6 displays the percentage of classes that emerged in the new domains plotted as a function of training condition. Percentages were used because a different number of classes could emerge after training with one or two domains. Specifically, up to 10 classes could emerge in the tests that followed training with one domain (two potential classes in each of five domains), and up to eight classes could emerge in the tests that followed training with two domains (two potential classes in each of four domains).

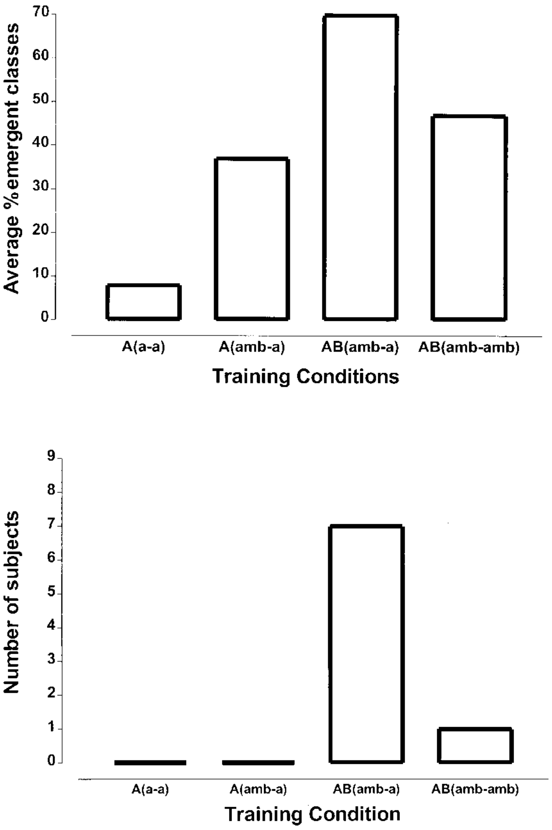


Fig. 6. Top: mean percentage of classes that emerged in the test domains, plotted as a function of training condition. Bottom: number of subjects who showed the emergence of at least 75% of the potential classes in the test domains, plotted as a function of training condition.

After A(a-a) training, few classes emerged in the new domains. After A(amb-a), AB(amb-a), and AB(amb-amb) training, however, a much larger percentage of classes emerged in the test domains. Thus, an increase in the number of domains, samples, or comparisons used for training resulted in an increase in the percentage of classes that emerged in the test domains.

A comparison of the results of A(a-a) and A(amb-a) training showed that the percentage of classes that emerged in the new domains was higher when many samples were used in training (Newman–Keuls post hoc pairwise comparison test, $q = 13.1$, $p < .001$). A comparison of the results of A(amb-a) and AB(amb-a) training showed that the percentage of classes that emerged in the new domains was higher when many domains were used in training ($q = 6.96$, $p < .001$). Indeed, AB(amb-a) training produced the highest level of spontaneous categorization of stimuli in new domains in this experiment.

A comparison of the results of AB(amb-a) and AB(amb-amb) training showed that the percentage of classes that emerged in the test domains was lower when many comparison pairs were used in training ($q = 4.71$, $p < .001$). The percentage of classes that emerged after AB(amb-amb) training was similar to that obtained after A(amb-a) training ($q = 2.25$, $p > .05$). Thus, the inclusion of many comparison pairs appeared to neutralize the enhancing effects of multiple-domain training on the emergence of a generalized categorization repertoire.

Generalized categorization in individuals. The presence of a generalized categorization repertoire in an individual can be defined as the emergence of at least a given percentage of the potential classes in the test domains. In the present experiment, the presence of a generalized categorization repertoire in an individual was defined by the emergence of at least 75% of the possible classes in the test domains. The lower panel of Figure 6 shows how each training condition influenced the number of subjects who showed the emergence of the generalized categorization repertoire.

After A(a-a), A(amb-a), or AB(amb-amb) training, few to none of the subjects showed the emergence of at least 75% of classes in the new domains. After AB(amb-a) training, however, most subjects showed the emergence of at least 75% of the classes in the test domains. When the results of AB(amb-a) training were compared to the outcomes of the other three conditions combined, the difference was significant (Fisher's exact test, $p < .0001$). Thus, the only training procedure that induced a generalized categorization repertoire was the exclusive combination of training with many samples in more than one stimulus domain. When compared to AB(amb-a) training, the addition of many comparison pairs in the AB(amb-amb) condition resulted in a very large and significant decrement in the number of subjects who showed the emergence of at least 75% of the classes in the new domains (Fisher's exact test, $p = .015$). Thus, the use of many comparison sets for training actually inhibited the enhancing effects of multiple-domain training.

Discriminability of stimuli in emergent classes. One of the criteria that must be satisfied to conclude that stimuli are functioning as members of a class is that many of the stimuli to which the response generalizes must be discriminable from each other (Cook et al., 1990; Fields & Reeve, 2001; Fields et al., 1997; Keller & Schoenfeld, 1950; Lea, 1984; Wasserman et al., 1988). If the stimuli are indistinguishable from each other, responding to them is functionally equivalent to responding that is under the control of a ''single'' stimulus. Thus, considering the stimuli as if they were a ''class'' would be essentially meaningless (Cook et al., 1990; Fields et al., 1997; Lea, 1984; Wasserman et al., 1988).

At least two approaches have been used to assess the discriminability of stimuli in a putative class. One is to infer discriminability among stimuli when a given response is occasioned with different likelihoods by different stimuli (Bhatt, Wasserman, Reynolds, & Knauss, 1988; Honig & Stewart, 1988; Lane, Clow, Innis, & Critchfield, 1998; Lea, 1984; Reeve & Fields, 2001). Another is to infer discriminability among stimuli when different stimuli occasion the same response but with different reaction times. Indeed, reaction time has been used by psychophysicists to measure detectability of pitch (Flynn, 1943), wavelength (Blough, 1978), and luminance (Raben, 1949); by cognitive psychologists to measure differential strength of relations among the stimuli in semantic memory networks (Balota & Lorch, 1986; Chiavello, Senehi, & Nuding, 1987; Collins & Quillian, 1969); and by behavioral psychologists to measure attention (Blough, 1993) and the strength of relations among stimuli in equivalence classes (Bentall, Jones, & Dickins,

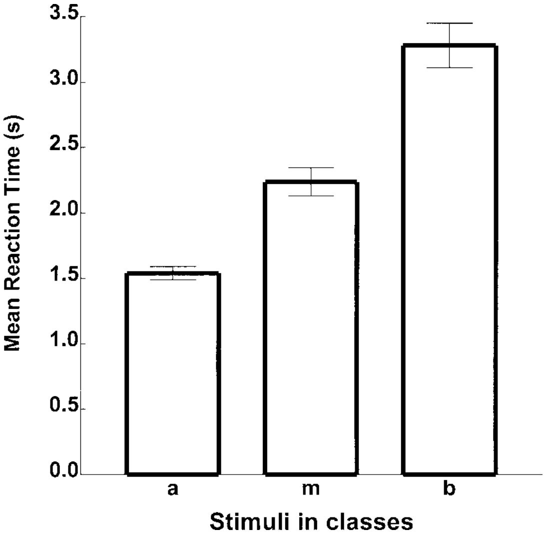


Fig. 7. Mean reaction times occasioned by the anchor, midpoint, and boundary stimuli in the emergent classes. Error bars show ±1 *SEM.*

1999; Fields, Landon-Jimenez, Buffington, & Adams, 1995; Spencer & Chase, 1996).

In the present experiment, reaction time was used to assess the discriminability of the stimuli in the emergent classes. Figure 7 displays the mean reaction times occasioned by the anchor, midpoint, and boundary stimuli in the classes that emerged in the test domains. The data were averaged across subjects, domains, and classes in a domain because no systematic differences in reaction time were found across these variables. The shortest average reaction time was occasioned by the anchor stimuli in a class. The reaction times occasioned by the midpoint and boundary stimuli increased as a direct function of their distance from the anchor stimulus. The results of an ANOVA showed the reaction times occasioned by the stimuli in classes to be significantly different from each other, $F(2, 579) = 54.19$, $p < .0001$. Newman–Keuls post hoc tests of pairwise comparisons showed significant differences in the reaction times occasioned by anchor and midpoint stimuli ($q = 5.86$, $p < .001$), the midpoint and boundary stimuli ($q = 8.77$, $p < .001$), and the anchor and boundary stimuli ($q = 14.63$, $p < .001$). Thus, the three stimuli in the emergent classes were discriminable from each other.

Failures of spontaneous categorization. Table 2 summarizes the effects of each training condition on the prevalence of those patterns of responding for the stimulus sets that did not occasion class-indicative responding. Identity conditional discriminations emerged in 96% to 100% of the tests after A(a-a), A(amb-a), or AB(amb-a) training. After AB(amb-amb) training, however, intact identity conditional discriminations emerged in only 79% of tests. This difference was significant (Fisher's exact test, $p = .0002$). Responding was not systematically related to stimulus value in 0% to 1% of the test sets presented after A(a-a), A(amb-a), or AB(amb-a) training and in 13% of the test sets presented after AB(amb-amb) training. This difference also was significant (Fisher's exact test, $p = .0006$). Thus, many-to-many training appeared to interfere with the emergence of (a) identity conditional discriminations in new domains and (b) dimensionial control of responding in new domains.

When classes did not emerge, in most cases performance in the generalization tests indicated that behavior was controlled by the dimensional value of the stimuli in a domain. The widths of the gradients, however, varied with training condition. A(a-a) training was followed by the emergence of many sharp gradients and relatively few moderate-width gradients. In contrast, A(amb-a), AB(amb-a),

Table 2

Percentage of generalization gradients measured in the new domains that occasioned different performance criteria when classes did not emerge.

Performance criterion	Training condition			
	A(a-a)	A(amb-a)	AB(amb-a)	AB(amb-amb)
Identity conditional discrimination	96	100	100	84
No dimensional control	1	4	0	16
Sharp gradients	80	32	23	13
Moderate gradients	19	65	77	71

or AB(amb-amb) training was followed by the emergence of many moderate-width gradients and relatively few sharp generalization gradients. All of the latter conditions contained at least one form of multiple-exemplar training. Therefore, it appears that at least one form of multiple-exemplar training induced moderate degrees of generalization among the stimuli in the sets in new domains, even when a generalized categorization repertoire did not emerge.

DISCUSSION

Class Formation

As noted in the introduction, the stimuli in a set function as members of a perceptual class when (a) all of the stimuli in a set must occasion a common response with similar high probabilities, (b) the stimuli in different sets must occasion different responses, and (c) many of the stimuli in a class must be discriminable from each other. The last criterion must be satisfied so that common responding to members of a set cannot be attributed to a generic failure discriminate among the stimuli in a set (Cook et al., 1990; Fields & Reeve, 2001; Keller & Schoenfeld, 1950; Lea, 1984; Reeve & Fields, 2001; Wasserman et al., 1988). In the present experiment, the selection-based performances occasioned by the stimuli in the test domains satisfied the first two criteria. The third criterion was satisfied by the chronometric data occasioned by the anchor, midpoint, and boundary stimuli that were presented to assess the emergence of classes in the new domains. Thus, the emergence of the classes and discriminability of the stimuli in those classes were measured at the same time.

Determinants of Generalized Categorization after AB(amb-amb) Training

As noted in the introduction, we expected AB(amb-amb) training to maximize the spontaneous categorization of stimuli in new domains because conditional discriminations were established among the anchor, midpoint, and boundary stimuli in each of four potential classes. Surprisingly, a smaller percentage of classes emerged in the new domains after AB(amb-amb) training than after AB(amb-a) training. AB(amb-amb) training also interfered with the emergence of (a) identity conditional discriminations among the anchor stimuli in the new domains and (b) dimensional control of responding by the stimuli in new domains. The induction of both of these stimulus control deficits, then, could have interfered with the emergence of a generalized categorization repertoire. According to this process-based analysis, the inhibitory effects of many-to-many training might be overridden by exposure to preexperimental procedures that would strengthen the emergence of identity conditional discriminations, dimensional control of responding, or both. Such an outcome would demonstrate the extent to which these two forms of stimulus control could influence the emergence of a generalized categorization repertoire.

The puzzling effects of many-to-many training are also amenable to experimental analysis. The formation of a generalized categorization repertoire was optimal after AB(amb-a) training but was far less likely after AB(amb-amb) training. One factor that differentiated the two procedures was that many conditional discriminations were established between one sample and many comparisons during AB(amb-amb) training but not during AB(amb-a) training. Therefore, the inclusion of one-to-many training appeared to inhibit the establishment of a generalized categorization repertoire. If this analysis is correct, the formation of a generalized categorization repertoire should be even less likely if AB(a-amb) or one-to-many training were to be conducted than after AB(amb-amb) or many-to-many training. Although the effect of AB(a-amb) training was not studied in the present experiment, the predicted outcome would confirm the supposition that one-to-many training was the factor in many-to-many training that was responsible for the inhibition of the emergence of a generalized categorization repertoire.

Determinants of Generalized Categorization after A(a-a) Training

A small percentage of classes emerged in the test domains in which identity conditional discriminations were established with stimuli in one domain [i.e., after A(a-a) training]. When presented with stimuli in the new domains, the preponderance of sharp generalization gradients showed that subjects responded in accordance with the prior history of identity conditional discrimination training. The lack of generalization to stimuli in new domains probably reflects a subject's prior history of reinforcement for identity-based selections in the presence of stimuli from many domains. Thus, the history of identity conditional discrimination training in the A(a-a) condition may well have served as a contingency-based instructional prompt to continue to respond on the basis of identity.

Determinants of Generalized Categorization after A(amb-a) Training

In the A(amb-a) condition, training was conducted with many exemplars as samples in one domain. Some subjects categorized a modest proportion of the stimuli in the new domains. In addition, when classes did not emerge, the majority of the test sets occasioned relatively wide generalization gradients. A comparison of the results of A(a-a) and A(amb-a) training conditions demonstrated two effects of training with many samples in one domain. First, it induced a modest generalized categorization repertoire. Second, when stimuli in the new domains were not categorized, broad ranges of stimuli in the new domains were treated as being related to each other.

These results of training probably reflect two processes. First, training with variants in a single domain maximized the predictability of reinforcement by sets of contiguous stimuli in a region of a domain. Second, this sort of training minimized the predictability of reinforcement by idiosyncratic features of any single stimulus in the domain. Thus, the variants in a region of a domain came to function as members of a perceptual class (Cook et al., 1990; Reeve & Fields, 2001; Tiemann & Markle, 1990; Wright et al., 1988). As such, the sample-based multiple-exemplar contingencies established a categorization repertoire. At the same time, because training occurred in the context of stimuli in only one domain, the categorization repertoire probably also came under the partial contextual control of the stimuli used in the training domain. Thus, only a modest degree of generalization of the categorization repertoire occurred to stimuli in new perceptually distinct domains.

It follows from this analysis that the categorization of stimuli in new domains in the present experiment should be directly related to their perceptual similarity to the stimuli in the A domain. This prediction can be evaluated by a reconsideration of the data collected. The categories in the A domain were constructed from naturally occurring stimuli: male and female faces. Domains C and E were also constructed of naturally occurring stimuli (trucks, cars, cats, and trees) that were perceptually similar to the stimuli used in the A domain. In contrast, Domains D and F were constructed of highly abstracted satellite imagery and thus were perceptually dissimilar to the stimuli used in the A domain. This analysis, then, would predict more spontaneous categorization of stimuli in Domains C and E than in Domains D and F. Of the 36 potential emergent classes in two domains (two potential classes in each of two domains for 9 subjects), 19 were categorized in the C and E domains and only 11 were categorized in the D and F domains. This difference was significant when evaluated with a directional chi square analysis, $\chi^2(1) = 4.00$, $p = .023$. This outcome, then, supports the contextual control analysis presented above.

Determinants of Generalized Categorization after AB(amb-a) Training

In the AB(amb-a) condition, training was conducted with stimuli in two perceptually distinct domains and with many exemplars presented as samples in each domain. Thereafter, many subjects categorized many of the stimuli in the new domains. Because more generalization of categorization occurred to stimuli in new domains after AB(amb-a) than A(amb-a) training, the increased categorization of stimuli in new domains could be attributed to training with stimuli in a number of domains.

The reinforcement of categorization-indicative responding in the presence of stimuli in two perceptually distinct training domains increased the range and variety of stimulus features that came to control that repertoire. Some of these features were likely to be similar to those found in the new domains. As such, the categorization repertoire generalized to a much broader range of new stimuli after multiple-domain than single-domain training (Fields & Reeve, 2000). It follows from this analysis that high levels of categorization should be occasioned by stimuli in domains that are similar to the A domain stimuli (i.e., the C and E domains). Likewise, high levels of categorization should be occasioned by stimuli in domains that are similar to the B domain stimuli (i.e., the D and F domains). This analysis would also predict that after training with stimuli in the A and B domains, the categorization of stimuli in the C and E domains should not differ significantly from the levels of categorization occasioned by the stimuli in the D and F domains. Of the 36 potential emergent classes in the C and E domains and the D and F domains (two potential classes in each of two domains for 9 subjects), 27 were categorized in the C and E domains and 23 were categorized in the D and F domains. This difference was not significant, $\chi^2(1) = 1.05$, $p = .31$. Therefore, after training with stimuli in two domains, the categorization repertoire became independent of the features of the stimuli in the training domains. Multiple-domain training, then, substantially attenuated the contextual constraints imposed on a generalized categorization repertoire.

To summarize, sample-based multiple-exemplar training with stimuli in one domain appears to be sufficient to establish a categorization repertoire. The reinforcement of categorization-indicative behavior in the presence of stimuli in perceptually distinct domains probably was responsible for generalization of the categorization repertoire to a wide range of perceptually distinct stimulus domains. That is, responding in the same manner to a set of physically similar yet discriminable stimuli—a behavioral unit called the categorization repertoire—itself comes under the control of stimuli in perceptually distinct stimulus domains, and thus is occasioned by stimuli in a wide range of new domains. Two predictions of this contextual control analysis were borne out with the data obtained in the present experiment.

A further evaluation of the theoretical analysis awaits new experiments that explicitly manipulate the perceptual similarity of domains used for training and testing. For example, after conducting B(amb-a) training, subjects should be more likely to categorize the stimuli in the D and F domains than in the C and E domains. In addition, training with stimuli in a few domains that show perceptual similarities ought to strengthen the categorization repertoire for stimuli in perceptually similar domains and increase discriminations between domains that are perceptually dissimilar. This analysis, then, would predict an increase in the categorization of stimuli in the C domain and a decrease in the categorization of stimuli in the D and F domains. Experiments such as these would further clarify how training with many exemplars in a domain, many domains, and domain similarities would each influence the induction of a generalized categorization repertoire.

Summary and Conclusions

The results of this experiment suggest four factors that influence the establishment of a generalized categorization repertoire. First, multiple-exemplar training with sample stimuli in one domain should be sufficient to establish a categorization repertoire. Second, training with many sample variants in a single domain also brings the categorization repertoire under the contextual control of the stimuli in the training domain. Third, transfer of the categorization repertoire to other domains should be a direct function of perceptual similarity between the test domains and that used for training. Fourth, multiple-domain training appears to be necessary to induce generalization of the categorization repertoire to stimuli in new domains.

Two general conclusions can be drawn from the data. First, the multiple-exemplar manipulation has markedly different effects on the emergence of performances. Those differences depend on the training variables to which the multiple-exemplar manipulation is applied. Thus, multiple-exemplar training cannot be viewed as a generic manipulation that has a singular effect on emergent performances.

Second, a generalized categorization repertoire is only one of many generalized repertoires that include learning set, identity matching, mirror-image and associative symmetry, and imitation. Given the common functional properties of all of these classes of behavior, the factors that influence the emergence of a generalized categorization repertoire might have similar effects on the emergence of these other complex repertoires that generalize to stimuli in new domains. The exploration of these variables should provide a strong procedural basis for understanding the emergence of complex generalized repertoires in natural settings.

REFERENCES

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 336–345.

Belanich, J., & Fields, L. (in press). Generalized equivalence classes as response transfer networks. *The Psychological Record.*

Bentall, R. P., Jones, R. M., & Dickins, D. W. (1999). Errors and response latencies as a function of nodal distance in five-member equivalence classes. *The Psychological Record, 49,* 93– 116.

Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., Jr., & Knauss, K. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 3,* 219–234.

Bhatt, R. S., & Wright, A. A. (1992). Concept learning by monkeys with video picture images and a touch screen. *Journal of the Experimental Analysis of Behavior, 57,* 219–226.

Blough, D. S. (1978). Reaction times of pigeons on a wavelength discrimination task. *Journal of the Experimental Analysis of Behavior, 30,* 163–167.

Blough, D. S. (1990). Form similarity and categorization in pigeon visual research. In M. L. Commons, R. J. Herrnstein, S. Kosslyn, & D. Mumford (Eds.), *Models of behavior: Behavioral approaches to pattern recognition and concept formation* (pp. 129–144). Hillsdale, NJ: Erlbaum.

Blough, D. S. (1993). Reaction time drifts identify objects of attention in pigeon visual search. *Journal of Experimental Psychology: Animal Behavior Processes, 19,* 107–120.

Boelens, H., & Van Den Broek, M. (2000). Influencing children's symmetric responding in matching-to-sample tasks. *The Psychological Record, 50,* 655–670.

Bourne, L. E., Dominowski, R. L., & Loftus, E. F. (1979). *Cognitive processes.* Englewood Cliffs, NJ: Prentice Hall.

Brown, A. K., Brown, J. L., & Poulson, C. L. (1995). Generalization of children's identity matching-to-sample performances to novel stimuli. *The Psychological Record, 45,* 29–43.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1965). *A study of thinking.* New York: Wiley.

Cerella, J. (1979). Visual classes and natural categories in the pigeon. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 68–77.

Chiavello, C., Senehi, J., & Nuding, S. (1987). Semantic priming with abstract and concrete words: Differential asymmetry may be post lexical. *Brain and Language, 31,* 43–60.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240–248.

Cook, R. G., Wright, A. A., & Kendrick, D. F. (1990). Visual categorization by pigeons. In M. L. Commons, R. J. Herrnstein, S. Kosslyn, & D. Mumford (Eds.), *Models of behavior: Behavioral approaches to pattern recognition and concept formation* (pp. 187–214). Hillsdale, NJ: Erlbaum.

Cross, D. V., & Lane, H. (1962). On the discriminative control of concurrent responses: The relations among response frequency, latency, and topography in auditory generalization. *Journal of the Experimental Analysis of Behavior, 5,* 487–496.

Delius, J. D., & Habers, G. (1978). Symmetry: Can pigeons conceptualize it? *Behavioral Biology, 22,* 336–342.

Dittrich, W. H., & Lea, S. E. G. (1993). Motion as a natural category for pigeons: Generalization and a feature positive effect. *Journal of the Experimental Analysis of Behavior, 59,* 115–130.

Dube, W. V., Iennaco, F. N., Rocco, F. J., Kledaras, J. B., & McIlvane, W. J. (1992). Microcomputer-based programmed instruction in identity matching to sample for persons with severe disabilities. *Journal of Behavioral Education, 2,* 29–51.

Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior, 53,* 345–358.

Fields, L., Landon-Jimenez, D. V., Buffington, D. M., & Adams, B. J. (1995). Maintained nodal distance effects after equivalence class formation. *Journal of the Experimental Analysis of Behavior, 64,* 129–146.

Fields, L., & Reeve, K. F. (2000). Synthesizing equivalence classes and natural categories from perceptual and relational classes. In J. C. Leslie & D. Blackman (Eds.), *Issues in experimental and applied analyses of human behavior* (pp. 59–84). Reno, NV: Context Press.

Fields, L., & Reeve, K. F. (2001). A methodological integration of generalized equivalence classes, natural categories, and cross modal perception. *The Psychological Record, 51,* 67–88.

Fields, L., Reeve, K. F., Adams, B. J., Brown, J. L., & Verhave, T. (1997). Predicting the extension of equivalence classes from primary generalization gradients: The merger of equivalence classes and perceptual classes. *Journal of the Experimental Analysis of Behavior, 68,* 67–92.

Figuracion, D. (1998). Morph, Version 2.5. San Diego, CA: Gryphon Software Corp.

Flynn, B. M. (1943). Pitch discrimination: The form of the psychometric function and simple reaction time to liminal differences. *Archives of Psychology (Columbia University).* No. 280 (p. 41).

Gena, A., Krantz, P. J., McClannahan, L. E., & Poulson, C. L. (1996). Training and generalization of affective behavior displayed by youth with autism. *Journal of Applied Behavior Analysis, 29,* 291–304.

Goldstein, H., & Mousetis, L. (1989). Generalized language learning by children with severe mental retardation: Effects of peers expressive modeling. *Journal of Applied Behavior Analysis, 22,* 245–259.

Handleman, J. S. (1979). Generalization by autistic-type children of verbal responses across settings. *Journal of Applied Behavior Analysis, 12,* 273–282.

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review, 56,* 51–65.

Hayes, S. C., & Barnes, D. (1997). Analyzing derived stimulus relations requires more than the concept of stimulus classes. *Journal of the Experimental Analysis of Behavior, 68,* 235–243.

Herrnstein, R. J. (1990). Levels of stimulus control: A functional approach. *Cognition, 37,* 133–166.

Herrnstein, R. J., & de Villiers, P. A. (1980). Fish as a natural category for people and pigeons. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 59–95). New York: Academic Press.

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 4,* 285–301.

Homa, D., & Chambliss, D. (1975). The relative contributions of common and distinctive information on the abstraction from ill-defined categories. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 351–359.

Homa, D., Cross, J., Cornell, D., Goldman, D., & Swartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology, 101,* 116–122.

Homa, D., & Little, J. (1985). The abstraction and long-term retention of ill-defined categories by children. *Bulletin of the Psychonomic Society, 23,* 325–328.

Homa, D., Sterling, S., & Treple, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 418–439.

Honig, W. K., & Stewart, K. E. (1988). Pigeons can discriminate locations presented in pictures. *Journal of the Experimental Analysis of Behavior, 50,* 541–551.

Hrycenko, O., & Harwood, D. W. (1980). Judgments of shape similarity in the Barbary dove (*Streptopelia risoria*). *Animal Behaviour, 28,* 586–592.

Hull, C. L. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs.* (Whole No. 123)

Innis, A., Lane, S. D., Miller, E. R., & Critchfield, T. S. (1998). Stimulus equivalence: Effects of a default-response option on emergence of untrained stimulus relations. *Journal of the Experimental Analysis of Behavior, 70,* 87–102.

Jitsumori, M. (1996). A prototype effect and categorization of polymorphous stimuli in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 22,* 405–419.

Keller, F. S., & Schoenfeld, W. N. (1950). *The principles of psychology.* New York: Appleton-Century-Crofts.

Lane, S. D., Clow, J. K., Innis, A., & Critchfield, T. S. (1998). Generalization of cross-modal stimulus equivalence classes: Operant processes as components in human category formation. *Journal of the Experimental Analysis of Behavior, 70,* 267–279.

Lea, S. E. G. (1984). In what sense do pigeons learn concepts? In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 263–276). Hillsdale, NJ: Erlbaum.

Lea, S. E. G., & Ryan, C. M. E. (1984). Feature analysis of pigeons' acquisition of concept discrimination. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 233–261). Hillsdale, NJ: Erlbaum.

Lubow, R. E. (1974). High-order concept formation in the pigeon. *Journal of the Experimental Analysis of Behavior, 21,* 475–483.

Malott, R., & Siddall, J. W. (1972). Acquisition of the people concept in the pigeon. *Psychological Reports, 31,* 3–13.

McIlvane, W. J., Dube, W. V., Green, G., & Serna, R. W. (1993). Programming conceptual and communication skill development. In A. P. Kaiser & D. B. Gray (Eds.), *Enhancing children's communication: Research foundations for intervention* (Vol. 2, pp. 243–285). Baltimore: Brookes.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Reviews of Psychology, 35,* 113–138.

Njegovan, M., Ito, S., Mewhort, D., & Weisman, R. (1995). Classification of frequencies into ranges by songbirds and humans. *Journal of Experimental Psychology: Animal Behavior Processes, 21,* 33–42.

O'Hora, D., Roche, B., Barnes-Holmes, D., & Smeets, P. (2002). Response latencies to multiple derived relations: Testing two predictions of relational frame theory. *The Psychological Record, 52,* 51–76.

Omohundro, J. (1981). Recognition vs. classification of ill-defined category exemplars. *Memory & Cognition, 9,* 324–331.

Pluchino, S. (1997). *The effects of multiple-exemplar training and stimulus variability on generalization.* Unpublished doctoral dissertation, Graduate Center of the City University of New York, New York.

Porter, D., & Neuringer, A. (1985). Music discrimination by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 138–148.

Raben, M. W. (1949). The white rat's discrimination of difference in intensity of illumination measured by a running response. *Journal of Comparative and Physiological Psychology, 42,* 254–272.

Reeve, K. F., & Fields, L. (2001). Perceptual classes established with forced-choice primary generalization tests and transfer of function. *Journal of the Experimental Analysis of Behavior, 76,* 95–114.

Rosch, E. H., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Smith, E. E. (1989). Concepts and induction. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 501–526). Cambridge, MA: MIT Press.

Spencer, T. J., & Chase, P. N. (1996). Speed analysis of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65,* 643–659.

Steele, D., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56,* 519–561.

Stokes, T. F., Baer, D. M., & Jackson, R. L. (1974). Programming generalization of a greeting response in four retarded children. *Journal of Applied Behavior Analysis, 7,* 599–610.

Tiemann, P. W., & Markle, S. M. (1990). *Analyzing instructional content: A guide to instruction and evaluation.* Champaign, IL: Stipes.

Wasserman, E. A., & DeVolder, C. L. (1993). Similarity and nonsimilarity-based conceptualization in children and pigeons. *The Psychological Record, 43,* 779–794.

Wasserman, E. A., Kiedinger, R. E., & Bhatt, R. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 3,* 235–246.

Wright, A. A., Cook, R. G., Rivera, J. J., Sands, S. F., & Delius, J. D. (1988). Concept learning by pigeons: Matching-to-sample with trial-unique video picture stimuli. *Animal Learning and Behavior, 16,* 436–444.

Young, J. M., Krantz, P. J., McClannahan, L. E., & Poulson, C. L. (1994). Generalized imitation and response-class formation in children with autism. *Journal of Applied Behavior Analysis, 27,* 685–697.

Received October 26, 2001
Final acceptance June 18, 2002

APPENDIX

To define the boundary, midpoint, and neither stimuli in all six of the domains, four graduate assistants who worked in the laboratory separately completed the following procedure. The unmorphed stimuli (the endpoints of each domain) were assigned morphing numbers 0 and 500, and were used as the anchor stimuli of potential Classes 1 and 2. Next, we printed colored pictures (2.5 cm by 2.5 cm) of 50 variants for each domain that were numbered by the morphing program in units of 10 (i.e., 0, 10, 20, . . ., 250, . . ., 480, 490, and 500). One anchor stimulus was placed on a table top and each laboratory assistant identified the stimulus that was most perceptually disparate from the anchor yet was still judged to be related to the anchor. This was accomplished by sorting the remaining 49 stimuli. This variant was labeled the boundary for that potential class.

To identify the midpoint stimulus in each class, the laboratory assistants were presented with the anchor and boundary stimuli from one potential class and were asked to select the stimulus that was perceptually equidistant from them. This was accomplished by sorting through all of the variants between the anchor and boundary for that potential class. The same procedure was then used to identify the anchor, midpoint, and boundary stimuli for the potential class at the other end of the domain. Finally, the boundary stimuli from the two potential classes were placed on the table, and the laboratory assistants selected a stimulus that was perceptually equidistant from them. This was accomplished by sorting through all of the variants between the boundary stimuli from the two potential classes. That variant was called the neither stimulus for that domain because it was a member of neither of the two potential classes.

Data collected from the laboratory assistants were averaged separately for each midpoint, boundary, or neither stimulus for each domain. The actual variants used as midpoint, boundary, and neither stimuli were the morphed image with numerical values that were closest to the computed averages for their corresponding stimuli. The values of the variants assigned to any midpoint, boundary, or neither stimulus varied by no more than two 10-point values across laboratory assistants. These data demonstrated the stability of ratings across laboratory assistants as well as the discriminability of the stimuli within and across classes in each domain.