JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1989, 52, 13-25 NUMBER 1 (JULY)

STIMULUS CLASS FORMATION AND CONCEPT LEARNING: ESTABLISHMENT OF WITHIN- AND BETWEEN-SET GENERALIZATION AND TRANSITIVE RELATIONSHIPS VIA CONDITIONAL DISCRIMINATION PROCEDURES

THOMAS G. HARING, CATHERINE G. BREEN, AND RICHARD E. LAITINEN

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

Three students with moderate mental retardation were taught a complex stimulus class with a two-choice conditional discrimination procedure applied across eight 10-member stimulus sets. Each set was composed of five age-appropriate and five age-inappropriate examples of clothing, accessories, and leisure items (e.g., a Walkman® radio). Discrimination training was programmed serially across each set, and generalization probes were conducted concurrently among all sets. Generalization probes consisted of unreinforced conditional matching trials with comparison items being drawn from (a) the set undergoing training (within-set probes), (b) sets not undergoing training (between-set probes), and (c) both sample and comparison items from different sets (transitive stimulus control probes). Results indicate that within-set generalization, between-set generalization, and transitive stimulus relations controlled responding by all 3 students for items that had been contingently associated with reinforcement. However, items that gained control of responding through within-set and between-set generalization alone (i.e., not acquired through contingent reinforcement) remained at baseline levels during transitive stimulus control probes. Results are discussed in terms of a taxonomy of multiple sources of stimulus control that underlie socially defined and maintained stimulus classes.

Key words: stimulus classes, concept learning, stimulus control, matching to sample, conditional discrimination, retarded youth

Demonstration of the formation of stimulus classes is an increasingly important focus of behavior-analytic research (e.g., Saunders, Wachter, & Spradlin, 1988). A stimulus class is defined as a generic class of interchangeable events that is related to a common response class (Skinner, 1935). The interchangeability of stimuli can be achieved through several known processes, such as the generalization of stimulus control across common stimulus features. The extent to which common stimulus features control responding, either jointly or in isolation, partially determines the boundaries of a generic stimulus class. For example, some stimuli are interchangeable only if they share all the essential features of the class, whereas others are interchangeable if they share some, but not all, of these features (see Wittgenstein, 1953). Stimuli can also become interchangeable if they occasion an equivalent functional effect or control a common response topography (such as when a child brandishes a stick, plastic baseball bat, or broom handle as a mock sword). In addition, it has been shown that physically dissimilar stimuli may become interchangeable through the extension of trained stimulus–stimulus relations (Sidman & Tailby, 1982). The purpose of the present study is to demonstrate the formation of a superordinate, socially determined stimulus class following the establishment of multiple stimulus–stimulus relations based on (a) the generalization of stimulus control across stimuli with a high degree of physical similarity (i.e., within-set generalization), (b) the extension of stimulus control to untrained sets of stimuli with a moderate degree of physical similarity to the stimuli of the trained sets (i.e., between-set generalization), and (c) the establishment of derived stimulus relations between stimuli from stimulus sets that share few or

This research was supported by the U.S. Department of Education, Office of Special Education and Rehabilitation Services, Field-initiated Research Grant G008630127. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Special Education and an official endorsement should not be inferred. The authors thank Robert Horner, Richard Saunders, and David Wacker for their thoughtful comments on earlier versions of this manuscript. We also thank Audrey O'Neil and the Office of the Santa Barbara Superintendent of Schools for the administrative support for this research. Correspondence and requests for reprints can be sent to Thomas Haring, Graduate School of Education, University of California, Santa Barbara, California 93106.

no physical features in common (i.e., transitive stimulus control).

The present study investigated the formation and resultant structure of a superordinate stimulus class of untrained stimulus relations that resulted from conditional discrimination training across multiple sets of age-appropriate clothing, accessory, and leisure items. Unlike stimulus classes in which one or more stimulus features are shared in common across all members of the class (i.e., basic level classes), a superordinate stimulus class is often based on established or derived equivalencies among physically dissimilar stimuli, which may or may not share common stimulus features (i.e., shared features are not necessary; Hoffman, 1981).

The stimulus class "trees," for example, is composed of members that share a necessary set of relevant stimulus features that determine "treeness." In contrast, a superordinate class is often based on socially established and maintained stimulus–stimulus relationships that determine which stimuli are and which stimuli are not members of the class. For example, the class of stimuli comprising appropriate social greetings can include such dissimilar events as a wave of the hand, a nod of the head, or saying "Hi." As individual members of a stimulus class these events share very few physical features in common, none of which is necessary, and none of which are inherently tied to the function of greeting someone.

In the present investigation we sought to teach the superordinate stimulus class of age-appropriateness to 3 adolescent girls with moderate and severe disabilities. The goal of training was to establish discriminative stimulus control by pictured exemplars and sets of pictured exemplars that defined a superordinate class of age-appropriate items. Within a two-choice conditional discrimination paradigm, participants were taught to select an age-appropriate exemplar in the presence of a star (*) and to select an an age-inappropriate exemplar in the presence of a square (□). While training was being programmed across stimulus sets in a sequential manner, the formation of a broader stimulus class was assessed by conducting generalization probes to monitor the establishment of discriminative stimulus control across untrained and unfamiliar age-appropriate and age-inappropriate exemplars (transitive stimulus control generalization), untrained members within sets of trained exemplars (within-set generalization), and exemplars from untrained sets of stimuli (between-set generalization).

Within-set generalization is defined as the extension of a response from a subset of trained stimuli to other untrained stimuli from a set whose members share all essential stimulus features in common. Thus, for a set comprised of examples *, A, B, and C (where the * designates the sample, and A, B, and C represent three items from a set such as shirts), if the relations *A and *B are taught, within-set generalization occurs if the subject responds correctly to *C in the absence of direct instruction or reinforcement. Engelmann and Carnine (1982) have developed an empirically validated model that explains within-set generalization on the basis of shared essential and nonessential stimulus features across the members of the set. The promotion of within-set generalization through multiple exemplar training is a widely replicated effect (e.g., Bellamy, Horner, & Inman, 1979; Gaylord-Ross, Haring, Breen, & Pitts-Conway, 1984; Guess, Sailor, Rutherford, & Baer, 1968).

Between-set generalization is defined as the extension of responding from a stimulus set established through direct training to a stimulus set that has not been directly trained or programmed. For example, if a student is taught to select age-appropriate shirts, between-set generalization occurs if responding is extended to age-appropriate shoes. Thus, for the stimulus classes *, A1, B1, C1 and *, A2, B2, and C2 (where the * designates a common sample stimulus and the remaining stimuli in the set consist of three positive comparison exemplars such as three shirts or three shoes), between-set generalization is demonstrated if, following conditional training on the relations *A1, *B1, and *C1, the student is able to respond appropriately to the relations *A2, *B2, and *C2 in the absence of direct instruction or reinforcement within that set. Theoretically, correct responding to the untrained set could occur because of shared stimulus features, shared functional effects, or shared response topographies across the sets. In the present research, between-set generalization could occur both because of stimulus features that are shared across sets and through the generalization of derived stimulus–stimulus relations established within the trained sets

(that is, the star signifies an age-appropriate choice).

Between-set generalization has received limited experimental evaluation. Garcia, Baer, and Firestone (1971) demonstrated that little between-class generalization occurred in the learning of generalized imitation by 4 retarded children. Parsonson and Baer (1978) assessed the occurrence of between-class generalization in the learning of generalized improvisation of tool use by preschool children. Haring (1985) provided a demonstration of between-class generalization by training retarded students to play with multiple sets of toys. After one example was trained from each class (e.g., one toy car, one toy airplane, etc.) generalization training was begun on toys from half of the sets. Concurrently, generalization probes were conducted between sets to the other half of the toy sets (to which only one example was trained). After students began to generalize within some sets as a result of being trained to generalize, between-set generalization emerged to sets where no generalization training was programmed.

Transitive stimulus generalization is defined as the extension of responding from directly taught stimulus–stimulus relations to stimulus–stimulus relations that are occasioned by a common intermediary stimulus (e.g., Sidman & Tailby, 1982). Thus, for the set A, B, C, and D, if AB, BC, and CD are taught, transitive generalization occurs if the subject responds correctly to AC, AD, and BD. Transitive stimulus generalization differs from within-set generalization because correct responding is controlled by the intermediary stimulus–stimulus relations established by training, and not on the basis of shared stimulus features across sets.

In the present study, transitive stimulus generalization probes were conducted by assessing derived stimulus–stimulus relations across sets. For example, if a student met criterion on shirts, pants, and hairstyles, a transitive generalization probe consisted of displaying an age-appropriate item as a sample (e.g., a shirt) along with one age-appropriate and one age-inappropriate item (e.g., a pair of pants and a hairstyle, respectively) as comparison stimuli. The student's task was to match the age-appropriate sample to the age-appropriate comparison. The relationship is potentially a transitive relationship because of the establishment of a common intermediary relationship for each stimulus; that is, the sets could become interrelated due to training with the common conditional stimulus (*), which may act as a node that links the previously unpaired stimuli (see Fields, Verhave, & Fath, 1984).

Spontaneous extension of responding to untrained items (within-set generalization), untrained items from untrained sets (between-set generalization), and untrained relations between stimuli (transitive relation generalization) indicates the formation of a conceptual stimulus class from originally unrelated sets of stimuli. The ability to form superordinate stimulus classes is a characteristic of human responding that may account for the comparatively rapid acquisition of complex repertoires with relatively little direct reinforcement. The complex social class of age-appropriateness is comprised of a system of interrelated stimulus classes that must ultimately come under the control of a generalized discriminative ability to select stimuli appropriately from multiple sets of exemplars. Studying the acquisition of this class provides an accessible experimental system for the study of between-set generalization, within-set generalization, and transitive stimulus control generalization. The present research was designed to (a) provide a preliminary attempt at demonstrating the formation of a complex stimulus class that comprises a social concept and (b) compare the emergence of within-set generalization to between-set generalization.

METHOD

Participants

Three female students in community-based classrooms for learners with moderate to severe handicaps participated in the study. Gina and Cherie both attended a self-contained class on a regular junior high school campus that emphasized instruction of functional daily living, vocational, and social skills. Both participants had numerous opportunities for interaction with nonhandicapped peers during class, lunch, and afternoon leisure periods, integrated physical education, cooking and sewing classes, and off-campus jobs.

Gina was 14 years old, was classified as moderately retarded with a Leiter IQ equivalent of 50, and was diagnosed as having Down

syndrome. She typically dressed in skirts or dresses with age-inappropriate graphics and wore black party shoes with bows. She often wore hair clips or bows detailed with age-inappropriate graphics (e.g., Cabbage Patch® dolls).

Cherie was 11 years old and was diagnosed as having moderate retardation with an IQ estimate of 58. Her receptive language ability was estimated to be an age-equivalent of 5 years, 9 months with the Peabody Picture Vocabulary test (revised edition). Cherie typically wore pants of polyester material that reached mid-calf level, t-shirts or sweatshirts emblazoned with age-inappropriate graphics, and thick-soled shoes that were highly atypical of shoe styles worn by her nonhandicapped same-age peers.

For 3 months prior to the introduction of the current research, Gina and Cherie were exposed to a daily program of appearance training by their classroom teacher that included whole group and individual discussions of age-appropriate clothing, leisure items, activities, and behavior. They also received daily feedback from their teacher and peers regarding their clothing selections. This combination of interventions showed no effect on either participant's selection of clothes or leisure objects at the time of the current investigation. Consequently, Gina and Cherie were identified by their classroom teacher as in greatest need of behavior change.

The third student, Vanessa, was 16 years old and attended a public high school. She was classified as moderately retarded with a Leiter IQ equivalent of 45, and was receiving an instructional program designed to teach community skills and domestic and self-care skills. Vanessa was integrated with nonhandicapped peers during breaks between classes, lunch periods, in her own class with peer tutors, and in on-campus jobs. Vanessa typically wore multiple t-shirts and sweatshirts (three or four) and rotated the one worn against her body on the previous day so that it became the current outer layer. The shirts were generally decorated with graphics that were either age-inappropriate or stigmatizing (e.g., with Special Olympics logos and slogans). Vanessa typically wore undersized pants and brown or blue thick-soled tennis shoes.

Vanessa had received appearance and age-appropriate behavior training in her classroom for 6 months prior to the beginning of the study. Intervention consisted of self- and teacher evaluations and feedback on a daily behavior change chart that assessed the presence or absence of appropriate clothing items and appropriate social behavior. Vanessa and her teacher discussed what determined age-appropriateness following each evaluation. Vanessa earned reinforcers contingent on age-appropriate clothing and behavior selection. During 6 months of intervention Vanessa showed no change in either her clothing and leisure activity selection or her age-appropriate social behavior.

All 3 students could identify younger children from students their own age. Pretests also demonstrated that all 3 students were capable of correctly labeling each item from the training sets. However, when asked to sort random pairs of age-appropriate and age-inappropriate items into piles representing "clothes that friends from other classes might wear to school" or "leisure items that friends from other classes might have," the students performed at chance levels.

Training-Stimulus Selection Procedures

An initial pool of 300 stimuli consisting of photographs of items taken from popular magazines and catalogs was developed. Twenty same-age nonhandicapped peers were then asked to rate the age-appropriateness of each item. Each peer was individually shown slides of each item and asked to rate its age-appropriateness on a 5-point Likert scale, with 1 indicating most appropriate and 5 indicating least appropriate items. All items that received an approval rating of 1 or 2 by at least 80% of the peers were included as positive training stimuli. Items that received an approval rating of 4 or 5 by at least 80% of the peers were included as negative training stimuli.

Eight sets of training stimuli consisting of pants, accessories, shoes, shirts, magazine covers, hairstyles, record covers, and sweaters were identified. Each set consisted of five age-appropriate and five age-inappropriate items. To the maximum extent possible, the sets were constructed of items that demonstrated the full range of stimulus features that defined both central and peripheral boundaries of the stimulus class (see Engelmann & Carnine, 1982; Horner, Sprague, & Wilcox, 1982). For example, items within the class of age-appro-

Table 1
Common characteristics for items in each set.

Sets	Characteristics of age-appropriate items	Characteristics of age-inappropriate items
Pants	Cotton material Pegged leg openings Zipper	Polyester material Flared leg openings Elastic waist
Shirts	Age-appropriate graphics Cotton material Small collar	Bows Polyester material Large collars
Sweaters	Age-appropriate graphics (different graphics than shirts) Long sleeves	Ruffles Age-inappropriate graphics
Shoes	Tennis shoes Thin soles White or pastel	Thick soles Brown or black
Magazines	Teenagers on cover	Children on cover
Records	Age-appropriate graphics (different graphics than sweaters) Young adults on cover	Age-inappropriate graphics (different graphics than sweaters) Older adults on cover
Accessories	Things a junior high student would take to school (e.g., radio, purse, backpack)	Things a child would take to school (e.g., toys, dolls, plastic preschool radio)
Hair	Hair worn loose Neatly combed	Ponytails Messy

priate pants shared such common features as cotton material, straight or tapering legs, and a zippered fly. Similarly, items within the class of age-inappropriate pants shared such common features as polyester material, floral patterns, and wide or bell-bottom legs. Sets were also constructed so that irrelevant properties and features (e.g., color) were randomly distributed across both positive and negative exemplars. Table 1 lists common features for items within each set.

Discrimination Training Procedures

Baseline. All training was conducted by the second author. The student was seated across from the instructor at a table (170 cm by 75 cm). A two-choice conditional match-to-sample format used. At the beginning of each trial the teacher placed one sample and two comparison pictures in front of the student. The comparison pictures were placed directly in front of the student, and the sample picture (either a star or square) was placed 10 cm above and directly between the two comparison pictures. The student responded by pointing to one of the comparison pictures. During baseline sessions, no feedback, rewards, prompting, or explanations of the task were given. Students were simply seated at the table and told to look at the sample symbol and choose a comparison picture. Two 40-trial sessions, with a 10 min break between each session, were conducted each day.

Conditional match-to-sample training. After stable baseline data were obtained, training was introduced for the first item of the first set (e.g., age-appropriate Shirt 1). During each training trial, the age-inappropriate comparison picture was randomly selected from among the five examples for that set (e.g., age-inappropriate shirts), and the age-appropriate example stayed the same. Correct responding was reinforced with praise on a variable-ratio (VR) 4 schedule; feedback ("yes" or "no") was given on every training trial. The criterion for acquisition of each positive exemplar was eight correct responses in a row. Switching from one training set to another occurred when responding to all age-appropriate items within the set met criterion, either through direct training or through generalization. A table of random numbers was used to determine the sequential order of the training sets for each participant. Cherie was trained with sweaters, album covers, shoes, shirts, accessories, pants, hairstyles, and magazines; Gina was trained with sweat-

ers, accessories, pants, record covers, shirts, shoes, magazines, and hairstyles; and Vanessa was trained with pants, shirts, accessories, hairstyles, shoes, magazines, record covers, and sweaters.

Generalization Probes

Five types of generalization probes were conducted throughout the study. These probes were identical to baseline procedures, with one to three probe trials interspersed every third or fourth training trial.

Within-set generalization probes. While examples from a set were being trained, within-set generalization to the remaining untrained examples of the set were probed on a daily basis. Untrained stimuli were presented with a randomly selected negative-set comparison stimulus.

Between-set generalization probes. Between-set generalization was probed by checking responding to examples from sets that had not yet received training. Again, a same-class, negative set comparison stimulus was contrasted to each positive exemplar probe trial.

Catalog rating probes. Each student was given a Sears catalog and told to select an outfit that consisted of a pair of pants, a shirt, a sweater, and a pair of shoes. These probes were conducted once during baseline and four to six times during intervention. Observers rated items chosen for the outfit according to the 5-point Likert scale used to select the training stimuli.

Shopping trip probes. Following training on each set of items, students were taken to a large department store and given the instruction to pick out an outfit (a pair of pants, a shirt, a sweater, and a pair of shoes) from the junior and shoe departments. The appropriateness of each item in the outfit was determined by the Likert scale procedure.

Transitive stimulus control probes. Two types of probes for transitive stimulus control were conducted during one baseline and over four to eight intervention sessions. These consisted of trained item and untrained item probes. Probes were conducted by using either an age-appropriate or age-inappropriate item in place of the star or square that had served as the sample stimulus during training. The two comparison stimuli consisted of an age-appropriate and age-inappropriate item from different stimulus sets. Trained item probes included items that had been directly associated with reinforcement during training (although their selection was never reinforced during transitive probes). Untrained item probes were conducted with items that met the criterion through either within- or between-class generalization; thus, selection of these items had never been reinforced.

Stimulus Control Validation Procedures

Validation of the basis of stimulus generalization across stimulus sets was conducted by having 10 adults, ages 19 to 43, identify common stimulus features that were shared among items from the pictured classes. Each participant was shown the entire array of training pictures and asked to identify common visual features that are shared among items from within and across the sets of pants, shirts, magazines, and so on. This task was facilitated by having the participant initially view two classes, with subsequent classes layered in one at a time until all of the training classes were displayed on a table. Each participant was also asked to state verbally the reason for their selections (e.g., "these pictures show bare arms and shoulders"). Data from this procedure were analyzed in terms of the interrelation of stimulus classes based on shared stimulus features that could mediate stimulus generalization across two or more stimulus classes. For example, several of the respondents identified "bare arms and shoulders" as a discriminable feature shared in common among two album cover and two magazine cover pictures.

Experimental Design, Measurement, and Reliability

A multiple probe variation of the multiple baseline across stimuli design was employed. Training was begun with the first item of the first set after stable baseline responding was obtained for all items across all sets. When responding to the first set of items met criterion (either through training or generalization), training within the second set was begun. Subsequent sets were introduced serially for training after responding to each item in the current training set met criterion. Concurrent generalization probes were interspersed throughout all training sessions.

Interobserver agreement checks were conducted every fifth session. The percentage of interobserver agreement was calculated ac-

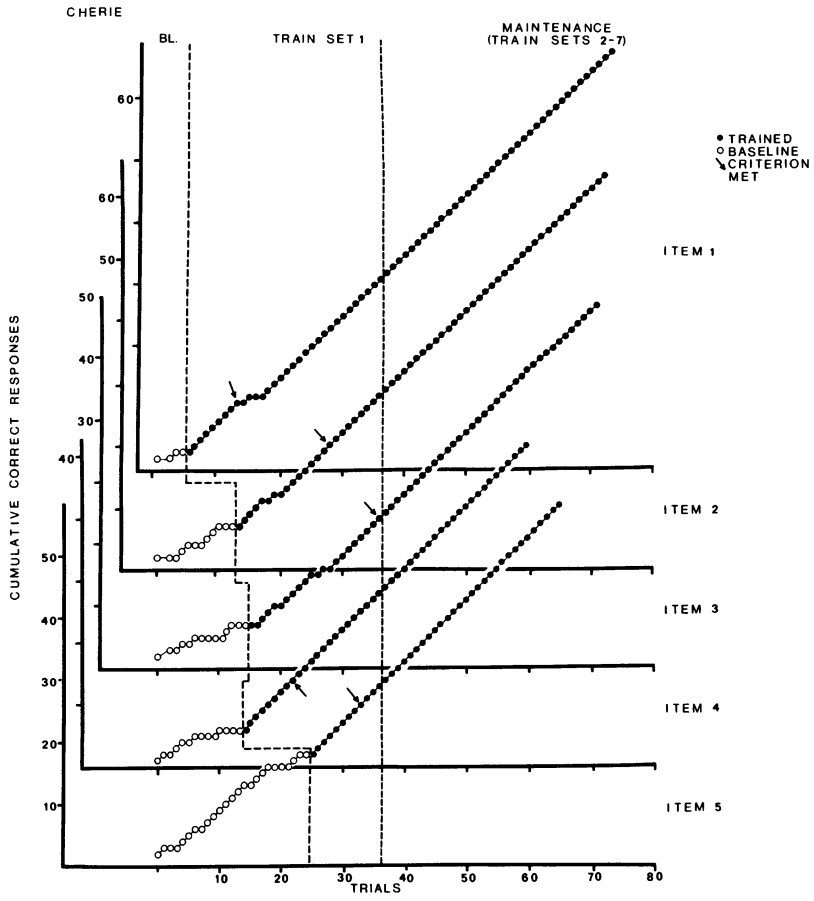

Fig. 1. The cumulative number of correct responses for Cherie's first set. All five examples required training. The baseline phase change line indicates the point at which training began on each item.

cording to the point-by-point correspondence method (Kazdin, 1982). Interobserver agreement data were collected across each phase and for each student. Interobserver agreement for baseline and training sessions was 100% on every session. Interobserver agreement for the appropriateness of students' clothing selections from catalog and store probes was assessed across 75 item selections. Of these 75 selections, the two raters had exact agreement of the 5-point scale on 32% of the items. The correlation between the two observers was .71.

RESULTS

Discrimination Training

Because training data were collected and analyzed for eight sets, with five items per set, there are 40 graphs for the positive stimuli for each participant. Due to the extensive nature of these data, we have selected three representative sets for display. The graphs were selected to show an initial set, a middle set, and the last set from the eight sets used with Cherie. Figure 1 shows the data for Cherie from the first set trained (sweaters). The data show that as training began with each item, the training criterion (eight consecutive correct responses) was rapidly met. For Items 1, 4, and 5, Cherie made only one error on the first trial for each item. For Item 2, she made three errors prior to reaching criterion, and for Item 3 she made five errors. In summary, Figure 1 shows that as training was layered in across items, the increase in correct responding was immediate and stable over the duration of the study.

Figure 2 shows Cherie's data from Set 5

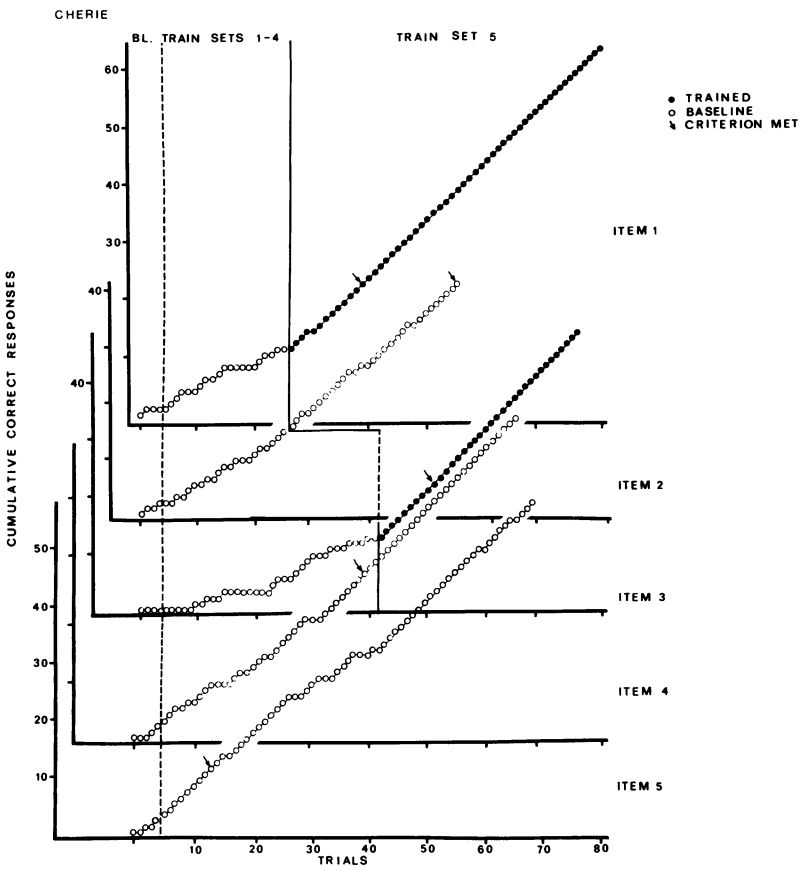

Fig. 2. The cumulative number of correct responses for Cherie's fifth set. Items 1 and 3 required training to reach the criterion, Items 2 and 4 reached the criterion while training was underway for the set (within-set generalization), and Item 5 met criterion during training on a preceding set (between-set generalization). The baseline phase change line indicates the point at which training began on Sets 1-4.

(pants). These data show that Item 1 and Item 3 required training. For Item 1, two errors were made, and for Item 3, one error was made prior to reaching criterion. Items 2, 4, and 5 reached criterion without receiving reinforcement. Items 2 and 4 met criterion after training was introduced for that set; therefore, these two items were classified as meeting the criterion for within-set generalization. Item 5 met criterion prior to the introduction of training for the set; therefore, this item was classified as meeting the criterion for between-set generalization.

Figure 3 shows Cherie's performance on the eighth set, magazines. The figure shows that training was not necessary on the eighth set. For four items (1, 2, 3, and 4) the criterion for acquisition was met after training was conducted across four prior sets (sweaters, record covers, shoes, and pants). Item 5 met criterion during the training of the fifth set (accessories). Because these items met criterion prior to training being conducted within this set, all five items were classified as meeting the criterion for between-set generalization

Within- and Between-Set Generalization

Figure 4 shows the cumulative number of items per set that met criterion for either within-set or between-set generalization for the 3 participants. Cherie's data are shown in the top panel of Figure 4, and indicate within-set generalization for nine items across the eight sets and between-set generalization for six items. Thus, of the 40 positive examples Cherie generalized (both within- and between-set) to 15 items (37.5%). Gina's data are represented in the middle panel of Figure 4, and show that

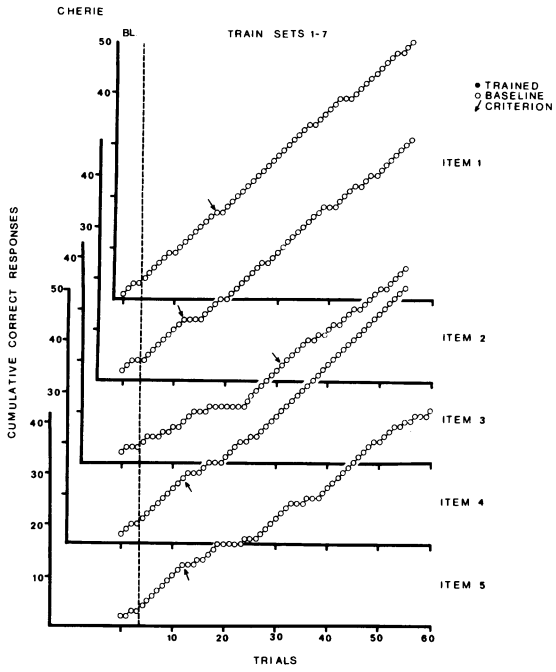

Fig. 3. The cumulative number of correct responses for the eighth set. None of the items from the eighth set required training, therefore all five items were classified as having achieved between-set generalization.

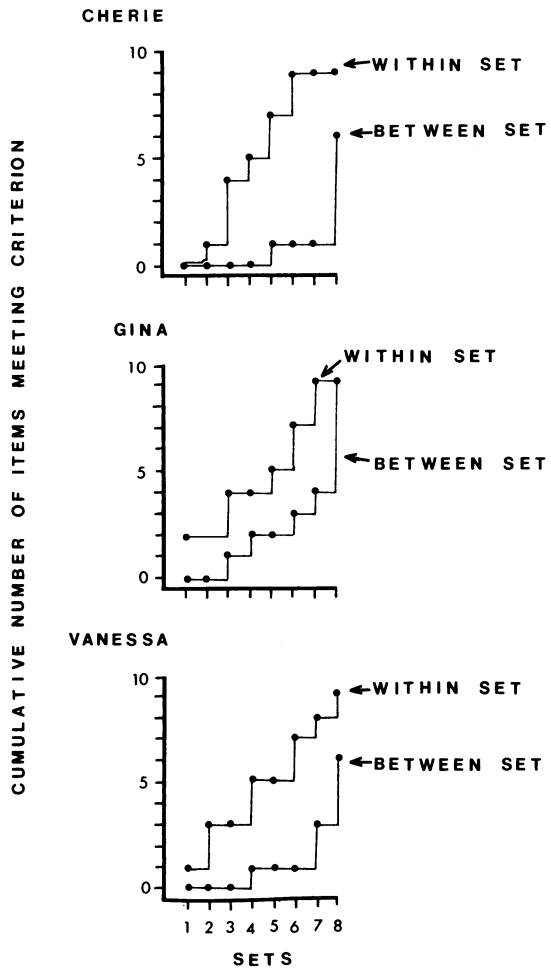

Fig. 4. The cumulative number of items that met the criterion for acquisition without training. Items meeting the criterion while other items in the set were being trained are labeled WITHIN-SET and items that met the criterion prior to other items in the set being trained are labeled BETWEEN-SET. Once items had generalized they were no longer included in subsequent probe trials.

Gina demonstrated within-set generalization to nine items and between-set generalization to nine items. Thus, Gina demonstrated generalization to 45% of the positive items. Vanessa's data are shown in the bottom panel of Figure 4. These data show that she demonstrated within-set generalization to nine items and between-set generalization to six items (37.5% of the items). In summary, all 3 subjects demonstrated both within- and between-set generalization. In addition, for all 3 subjects the pattern of generalization was similar in that within-set generalization occurred prior to between-set generalization, with between-set generalization requiring between three and five training sets to appear. By the final set, both Cherie and Gina demonstrated between-set generalization to all items, and Vanessa demonstrated between-set generalization to three of five items.

Transitive Stimulus Generalization Probes and Related Tasks

The data in the bottom panel of Table 2 represent the subjects' performance on transitive probes. All 3 subjects received 10 trials during baseline sessions. When probes were conducted with items that met the criterion for acquisition through training, the performance was significantly above baseline levels, $t(4) = 4.90$, $p < .01$. However, when items met the criterion for acquisition through generalization (rather than through reinforced trials) the performance was not significantly above baseline levels. The data from the catalog probes and the shopping probes are given in the upper panel in Table 2. These data show that conditional discrimination training across multiple sets of items was associated with more ap-

Table 2

Mean performance on related tasks to that used in training and on transitive-equivalence probes.

	Related tasks[a]				Transitive-equivalence probes[b]		
	Catalog probes		Shopping probes				
Subject	Baseline	Training	Baseline	Training	Baseline	Trained items (meeting criterion)	Untrained items (meeting criterion)
1	3.8	2.7	3.8	2.2	40%	68%	39%
2	4.8	3.4	5	2.6	50%	83%	54%
3	3.4	2.8	3.2	2.1	30%	75%	56%

[a] Mean clothing rating for phases from Likert scale data: 5 = highly age inappropriate, 1 = highly age appropriate.
[b] Mean percentage of correct performance for each condition.

propriate clothing selections across both related tasks.

Results of the feature analysis by adult observers are summarized in Figure 5, which shows the network of relations as identified by the observers. The related classes and the number of observers who nominated them as sharing overlapping stimulus features are indicted by arrows and intermediary proportions (e.g., 7/10). These data show that 10 of 10 observers identified overlapping stimulus features among the magazine and music (album cover) sets, and 9 of 10 observers identified overlapping stimulus feature among music and sweater items (e.g., graphic design consisting of squares, triangles, and spirals). Of the many overlapping stimulus features that were identified, those features shared among the magazine and music and the music and sweater stimulus sets were the most reliably and consistently mentioned.

The network of stimulus classes whose overlapping features were identified by 3 or more observers is shown in Figure 5. This network accounts for 81% of the observed between-class generalizations on the basis of common stimulus features shared among individual members of the different stimulus classes. The percentage of between-class generalizations accounted for by shared stimulus features is reduced to 75% if the criterion for class inclusion in the network is based on nomination by 5 or more observers—resulting in a stimulus class network that includes magazines, music album covers, sweaters, and pants.

DISCUSSION

The results indicate that the conditional discrimination procedures resulted in both within-class and between-class generalization across the 3 participants. An initial purpose of the study was to demonstrate the formation of a broad stimulus class that comprises a social concept. The inference that a social concept emerged as a result of training is difficult to establish because such generalization could be attributable to an interaction between individual learning history (e.g., prior exposure to related social concepts such as young vs. old) and exposure to experimental conditions. Correct responding to items from untrained sets was our major test of social conceptual responding. The results suggest that a socially important stimulus class emerged as a result of the conditional discrimination training—all 3 subjects demonstrated between-set generalization. Between-set responding is indicative of the formation of a superordinate stimulus class because the correct responding occurred to items from sets that did not participate directly in discrimination training contingencies.

Data from the network analysis indicate that between-class generalization among common stimulus features may have played a significant role in the emergence of the superordinate stimulus class. However, generalized responses that are left unaccounted for by feature-generalization processes require further explanation. A recent study by Saunders, Saunders, Kirby, and Spradlin (1988) reports that handicapped students reliably assign stimuli to classes even though the assignments go unreinforced. Thus, arbitrary assignment may represent a general process whereby humans form stimulus classes in the absence of reinforcement. Thus, some of the between-class generalization observed in this study may reflect a participant's arbitrary assignment of stimulus items to the "star" and "square"

stimulus classes during probe conditions. Other probe items may have been generalized to because they share features in common with stimuli in the student's daily environment. For example, generalizations to items from the hairstyle set may have been influenced by exposure to hairstyles worn by nonhandicapped peers. In addition, generalization to more remote stimulus situations (e.g., shopping and catalog probes) indicate the emergence of mediational relations such as generalized delayed match to sample.

For all 3 subjects, within-set generalization occurred prior to between-set generalization. Although this pattern was also identified by Haring (1985), it is not known whether this is due to the training procedure used (in which items were taught in a set-by-set manner) or to a more general process. Theoretically, this pattern of within-set and between-set generalization may indicate that, just as *multiple-exemplar* training functionally controls the occurrence of within-set generalization (e.g., Engelmann & Carnine, 1982), *multiple-set* training may control the occurrence of between-set responding. In the present study, subjects required training across a minimum of three sets (range, three to five sets) before between-set generalization was demonstrated. The theoretical interpretation that socially established and maintained stimulus classes are a function of training across multiple-stimulus sets is appealing because it permits an explanation of such responding on the basis of exposure to controllable learning processes (e.g., delayed matching, stimulus equivalence, stimulus generalization, etc.) rather than on unobservable congnitive processes.

An additional purpose of the study was to investigate the nature of the stimulus class that was formed. An important question in this regard is whether or not the class, as it was established, was isomorphic with an equivalence class as described by Sidman and Tailby (1982). To answer this question, we collected data on performance during transitive-equivalence probes. The results indicated that the subjects responded at levels that were significantly above baseline performance on transitive probes that involved trained items in novel relation to each other (i.e., a reversal of the sample–comparison relations established during training). However, with items that met criterion through generalization rather than

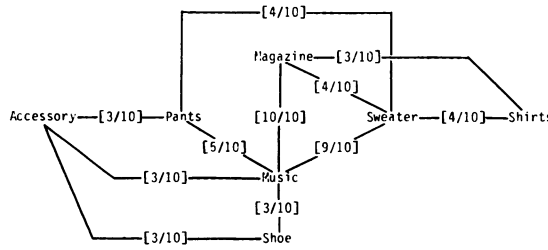

Fig. 5. Network of stimulus features shared in common across several stimulus classes. This network represents the composite reliability among 10 adults, with the bracketed proportions indicating the number who agreed on the presence of a common feature across classes.

direct training, the students performed at essentially chance levels. This indicates that the class, as it was formed, was not an equivalence class in that the untrained items, although they met criterion for acquisition, did not demonstrate generalized transitive stimulus equivalence.

A possible explanation for these findings is offered in the investigations of Fields et al. (1984) that identify important differences between trained and untrained transitive probes. For example, transitive stimulus control probes of untrained items (both within and across classes) differ from probes of trained items in terms of the schedules of reinforcement associated with each (i.e., an extinction schedule for untrained items, and a VR schedule for trained items), and the number of intermediary equivalency relations that must be established for the emergence of transitive control among untrained items (Sidman & Tailby, 1982). For example, transitive stimulus control among untrained items requires that equivalency relations be established during unreinforced presentations of the training sample stimulus (the star or square) with untrained comparison stimuli (e.g., during between-class probes), and that these relations maintain during unreinforced presentations of novel sample–comparison pairings of items from different sets. The influence on untrained transitive stimulus control by insertions of noncontingent novel (i.e., between-class probes) and noncontingent novel reversed (i.e., transitive stimulus control probes) sample–comparison trials into the training sequence suggests an area for future research.

The catalog and department store probe data showed that the students were capable of gen-

eralizing from the picture training (as conducted during the discrimination training phases) to the similar task of choosing appropriate pictures from a catalog, a task that shares many of the stimulus features and characteristics of the original training task, and the less similar task of choosing new clothing at a store, a task that requires the generalization of stimulus control from pictures to objects in an entirely different setting. The latter instance of generalization is interesting in that it indicates a modification of behavior embedded in an untrained chain of responses occurring within a significantly different environment under the control of remote training contingencies.

If generalization is viewed as a multivariate dependent variable consisting of separate subclasses of generalization within and between training and nontraining (probe) stimulus classes, as was attempted within this study, then additional theoretical and empirical questions arise. To study generalization as a multivariate dependent variable, a primary task is to define the subclasses of generalization in terms of their units of analysis (Johnston & Pennypacker, 1980). Thus, in the present research three subclasses of generalization were defined: within- and between-set generalization and transitive-equivalence generalization, with the former two being driven by feature generalization mechanisms and the latter by stimulus–stimulus generalization mechanisms (cf. Sidman, 1986). It is possible, however, that generalization may be under the control of a single variable, as Stokes and Baer (1977) have suggested. Thus, a question for future research is to determine whether generalization functions as a single unit or as a set of distinct subunits that interact in unique ways. By way of analogy, generalization may operate as a "large operant" similar to a tact, mand, or intraverbal (Skinner, 1957), or as a set of smaller subunits similar to subclasses of the intraverbal (Chase, Johnson, & Sulzer-Azaroff, 1985). Thus, in the present study it is possible that an interrelated subset of generalization operants (within-set, between-set, and transitive-equivalence probes with trained items) was established. A preliminary analysis of these data suggests that within- and between-set generalization may operate in a separate yet dependent manner in that within-set generalization consistently emerged prior to between-set generalization.

In summary, the present study adds to our knowledge of between-set generalization. The processes involved in this type of generalization are not yet fully understood, and data that show the effects of these processes on the formation of complex stimulus classes are lacking. Knowledge of these processes could affect substantially the organization of instruction for learners with moderate and severe handicaps and could lead to a more effective instructional technology whereby complex stimulus classes can be established through the planned control of generalization within and across stimulus classes. In addition, the study posits a plausible behavior-analytic explication of social-conceptual learning and responding.

REFERENCES

Bellamy, G. T., Horner, R. H., & Inman, D. P. (1979). *Vocational habilitation of severely retarded adults: A direct service technology.* Baltimore: University Park Press.

Chase, P. N., Johnson, K. R., & Sulzer-Azaroff, B. (1985). Verbal relations within instruction: Are there subclasses of the intraverbal? *Journal of the Experimental Analysis of Behavior,* **43,** 301–313.

Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications.* New York: Irvington.

Fields, L., Verhave, T., & Fath, S. (1984). Stimulus equivalence and transitive association: A methodological analysis. *Journal of the Experimental Analysis of Behavior,* **42,** 143–157.

Garcia, E., Baer, D. M., & Firestone, I. (1971). The development of generalized imitation within topographically determined boundaries. *Journal of Applied Behavior Analysis,* **4,** 101–112.

Gaylord-Ross, R. J., Haring, T. G., Breen, C., & Pitts-Conway, V. (1984). The training and generalization of social interaction skills with autistic youth. *Journal of Applied Behavior Analysis,* **17,** 229–247.

Guess, D., Sailor, W., Rutherford, G., & Baer, D. M. (1968). An experimental analysis of linguistic development: The productive use of the plural morpheme. *Journal of Applied Behavior Analysis,* **1,** 297–306.

Haring, T. G. (1985). Teaching between-class generalization of toy play behavior to handicapped children. *Journal of Applied Behavior Analysis,* **18,** 127–139.

Hoffman, M. L. (1981). Perspective on the difference between understanding people and understanding things: The role of affect. In J. H. Flavell & L. Ross (Eds.), *Social cognitive development: Frontiers and possible futures* (pp. 67–81). New York: Cambridge University Press.

Horner, R. H., Sprague, J., & Wilcox, B. (1982). General case programming for community activities. In B. Wilcox & G. T. Bellamy (Eds.), *Design of high school programs for severely handicapped students* (pp. 61–98). Baltimore: P. H. Brookes.

Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research.* Hillsdale, NJ: Erlbaum.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.* New York: Oxford University Press.

Parsonson, B. S., & Baer, D. M. (1978). Training generalized improvisation of tools by preschool children. *Journal of Applied Behavior Analysis,* **11,** 363–380.

Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Behavior,* **50,** 145–162.

Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior,* **49,** 95–115.

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.

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.

Skinner, B. F. (1935). The generic nature of the concepts of stimulus and response. *Journal of General Psychology,* **12,** 40–65.

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis,* **10,** 349–367.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). New York: Macmillan.

Received February 9, 1989
Final acceptance March 15, 1989