The Psychological Record, 2002, 52, 77-88

STIMULUS EQUIVALENCE AND NONARBITRARY RELATIONS

IAN STEWART, DERMOT BARNES-HOLMES, AND BRYAN ROCHE
National University of Ireland, Maynooth

PAUL M. SMEETS
Leiden University, Holland

This study investigated the effect of conflicting nonarbitrary (color) relations on equivalence responding. Three groups of 8 subjects were trained and tested for the formation of three 3-member equivalence classes using nonsense syllables as stimuli. Each subject received two separate exposures. For the No-Color group all stimuli in both training and testing phases were in black lettering. For the All-Color group, all stimuli were in color. Thus, training effectively involved learning to ignore color and, as predicted, during testing color had little or no effect upon performance, in that there was no significant difference in levels of equivalence responding between the no-color and all-color groups. For the Color-Test group, training and testing stimuli were in black and color lettering respectively. During testing, the sample was always differently colored from the "equivalent" comparison, but was the same color as one of the nonequivalent comparisons. These subjects had no history of reinforcement for ignoring color, and thus a possible conflict between arbitrary and nonarbitrary relational control was produced in this condition. Results showed that for this third group, levels of equivalence responding were significantly lower than for either of the other two groups. Furthermore, levels of responding in accordance with color matching were significantly higher for the color-test group than for the all-color group. These data are consistent with Relational Frame Theory.

All animals capable of complex forms of learning may be trained to make discriminations on the basis of nonarbitrary or physical relations between stimuli (e.g., louder than, larger than, differently colored from). However, according to Relational Frame Theory (RFT; Hayes, 1991), language-able humans, having prolonged exposure to certain contingencies of reinforcement operating within the verbal community, also demonstrate responding on the basis of derived or arbitrarily applicable relations. These relations are defined not by the physical properties of the relata per se, but by additional contextual cues (Hayes,

Requests for reprints should be addressed to Dermot Barnes-Holmes, Department of Psychology, National University of Ireland, Maynooth, Maynooth, Co. Kildare, Ireland.

Barnes-Holmes, & Roche, 2001; Hayes, 1991). If, for example, a language-able person is trained in a series of related conditional discriminations using a matching-to-sample format, often new and untrained relations among the stimuli involved in those discriminations will emerge. If the person is taught to select stimulus B in the presence of stimulus A and stimulus C in the presence of stimulus B then during subsequent testing he or she may well demonstrate several untrained "matching responses," including the following: the selection of A given B and of B given C (thus showing symmetry); the selection of C given A (thus showing transitivity), and the selection of A given C (thus showing a combination of symmetry and transitivity). The presence of these untaught or derived relations has been taken by behavior analysts as demonstrating that A, B, and C now participate in an equivalence class; a class of mutually substitutable elements (e.g., Barnes, Browne, Smeets, & Roche, 1995; Fields, Adams, Verhave, & Newman, 1990; Sidman, 1992). Equivalence class formation constitutes one example of arbitrarily applicable relational responding or relational framing (Barnes & Holmes, 1991; Hayes, 1991).

One prediction made by RFT is that there are certain conditions under which equivalence formation in verbally able human adults may be disrupted. One way in which this can happen is if one set of arbitrary relations comes into conflict with another set of arbitrary relations (e.g., Barnes, Lawlor, Smeets, & Roche, 1995; Watt, Keenan, Barnes, & Cairns, 1991). In the study by Watt et al., for example, it was shown that when the stimuli used during equivalence training participated in preexperimentally socially established equivalence relations, which conflicted with the experimenter-designated equivalence relations, equivalence responding was less likely to emerge. Relational Frame Theory also predicts that it should be possible to interfere with equivalence formation in verbally sophisticated adults when conflicting arbitrary and nonarbitrary stimulus relations are introduced into the relevant test procedures. More specifically, RFT makes a distinction between arbitrary and nonarbitrary relational responding, and it suggests that the latter is a likely precursor of the former (Barnes & Roche, 1996). Nonarbitrary relational responding[1] is likely established when a child learns to categorize formally similar objects together (e.g., red blocks in one pile and green blocks in another; Barnes & Roche, 1996). Although a history of nonarbitrary relational responding might be sufficient to produce reflexivity in a matching-to-sample context, it is unlikely to produce responding in accordance with symmetry and transitivity (i.e., arbitrarily applicable "sameness" responding). According to RFT, explicitly

[1]It is important to emphasize that nonarbitrary relational responding is learned behavior, but in this case the relational responses depend in part on formal relations between or among stimuli. The dimension of these formal relations may be selected arbitrarily by an experimenter, but the relational control itself is far from arbitrary. For example, in training identity matching along the dimensions of color, the experimenter *must* reinforce the matching of green to green rather than red to green.

reinforced symmetrical and transitive responding, such as that involved in learning the spoken, heard, and written names for objects and events in the world, plays an important role in the production of arbitrarily applicable sameness or equivalence responding. The key point here, is that although equivalence is based on a more extensive and arguably more complex reinforcement history than nonarbitrary relational responding, nonarbitrary sameness responding provides an important historical context for establishing its arbitrary counterpart (the relational frame of sameness). Indeed, it seems unlikely, except in the most artificial of learning contexts, that a child would demonstrate symmetry and transitivity without having previously acquired a repertoire of nonarbitrary sameness responding (see Hayes, 1991). Given this theoretical relationship between nonarbitrary and arbitrary sameness responding, there is considerable scope for investigating empirically the relationship between arbitrary and nonarbitrary relational responding. While some researchers have reported relevant findings in this area (Barnes & Keenan, 1993; Fields, Reeve, Adams, & Verhave, 1991), one issue that has not been examined is the extent to which conflicting nonarbitrary and arbitrary stimulus relations will interfere with one another. This was the purpose of the present study.

Method

Subjects

Twenty four students of University College, Cork acted as subjects. Of these subjects 19 were female and 5 were male and their ages ranged between 18 and 25 years. All subjects were volunteers who were contacted through personal acquaintances and chosen on the basis that they had no previous experience or knowledge of stimulus equivalence. Subjects were assigned randomly to one of three experimental conditions (i.e., 8 subjects per condition).

Apparatus

Subjects sat at a desk in a small experimental room, facing an Apple Macintosh LC - III computer and monitor. A single sheet of typed instructions was left on the desk. Three keys on the computer keyboard, Z, V, and M, were marked with white stickers; these were designated as response keys. Three different BBC BASIC computer programs (one per experimental condition) controlled the presentation of all relevant experimental stimuli as well as the recording of subjects' responses. The following nine nonsense syllables were used as experimental stimuli: ZID (A1), MAU (B1), JOM (C1), VEK (A2), WUG (B2), BIF (C2), YIM (A3), DAX (B3), PUK (C3). Henceforth, they will be referred to using their accompanying alphanumeric labels (subjects never saw these labels).

Experimental Overview

The current study set out to examine the extent to which subjects would respond on the basis of an arbitrary or nonarbitrary sameness relation when these two relations conflicted with one another. Imagine, for example, that a subject is presented with a green sample stimulus (C1) and three comparison stimuli (A1, colored red; A2, colored green; and A3, colored blue). If C1 and A1 are expected, based on training, to participate in a derived equivalence relation but are different colors, then subjects may respond either on the basis of an arbitrary sameness relation by choosing A1, or on the basis of the nonarbitrary relation by choosing A2, which is the same color as C1. To examine more closely the determinants of arbitrary or nonarbitrary relational responding in such a context, three groups of subjects were trained and tested for the formation of three 3-member equivalence classes. For Group 1, all stimuli were in black lettering against a white background, and thus no conflict was expected to occur between nonarbitrary and arbitrary relational control. For Group 2, stimuli during training and testing were in color. Moreover, during the training, reinforcement was delivered contingent upon responding towards a comparison that was the same color as the sample on some trials and responding towards a different color comparison on other trials. In effect, subjects were trained to "ignore" the nonarbitrary relation of color as a relevant dimension for relational responding. Thus, during testing, the color of the stimuli was expected to have little or no impact upon subjects' performances. In Group 3, however, subjects were trained with black stimuli before being tested with colored stimuli. In effect, subjects were not provided with a reinforcement history for ignoring the nonarbitrary relational dimension of color during training. In testing for this group, the correct (i.e., equivalent) comparison stimulus was always a different color from the sample, and one of the incorrect (i.e., nonequivalent) comparisons was always the same color as the sample. Thus, a conflict between nonarbitrary and arbitrary relational control was predicted. In short, RFT predicts that the color-test group should produce significantly lower levels of equivalence responding than the no-color and all-color groups.

Procedure

Each subject was exposed to two separate sessions of training and testing. Before starting each individual session, the experimenter informed the subject that he or she should read the typed instructions beside the computer keyboard and that a message appearing on the computer screen would signal the end of the experimental session. The typed instructions were as follows:

> During this experiment you will be presented with 3-letter nonsense syllables (e.g., CUG, DAX, VEK, etc.).
>
> You should look at the nonsense syllable that appears at the top of the screen, and then choose one of the nonsense syllables that appear below.
>
> In the first part of the experiment the computer will always tell you whether your choice was correct or wrong.

> In the latter part of the experiment, however, the computer will no longer provide you with feedback.
>
> You choose the nonsense syllable that appears on the left by pressing the marked key on the left.
>
> You choose the nonsense syllable that appears in the middle by pressing the marked key in the middle.
>
> You choose the nonsense syllable that appears on the right by pressing the marked key on the right.

On each matching-to-sample trial, one nonsense syllable, the sample stimulus, appeared in the top center of the screen and three other nonsense syllables, the comparison stimuli, appeared in a line along the lower half of the screen. Subjects were instructed to look at the top one (the sample) and then choose one of the three nonsense syllables (comparisons) appearing below by pressing the appropriate marked key. To choose the one on the left, subjects had to press the marked key (Z) on the left. To choose the one in the middle, they had to press the marked key (V) in the middle. To choose the one on the right, they had to press the marked key (M) on the right.

The experiment consisted of two phases: a training phase and a testing phase. During the training phase, the color format for experimental stimuli was as follows.

No-color group - Black lettering
All-color group - Color lettering
Color-test group - Black lettering

During this phase of the experiment, subjects were trained on three A-B and three B-C matching-to-sample tasks. For the three A-B tasks, subjects were presented with either A1, A2, or A3 as the sample stimulus and then had to choose from among the three comparison stimuli B1, B2, and B3. A correct response was B1 given A1, B2 given A2, and B3 given A3. For the three B-C tasks, subjects were presented with either B1, B2, or B3 as the sample stimulus and had to choose from among the 3 comparison stimuli C1, C2, and C3. A correct response was C1 given B1, C2 given B2, and C3 given B3. If a subject responded correctly, the stimulus display cleared and the word 'Correct' appeared in the center of the screen accompanied by a high-pitched beep for 1.5 seconds. If a subject responded incorrectly, the stimulus display cleared and the word 'Wrong' appeared in the center of the screen but without auditory feedback.

The matching-to-sample tasks were presented in a repeating cycle of 36 trials (see Appendix 1), the order of which was the same for every subject. First, the 3 A-B tasks were presented six times each in a quasi-randomly ordered block of 18 trials; the three B-C tasks were then presented six times each in another quasi-randomly ordered block of 18 trials. Across both these blocks, each of the following elements was counterbalanced: (a) the order of presentation of the 3 A-B, and then the 3 B-C matching-to-sample (MTS) tasks; (b) the spatial positioning of the comparison stimuli (left, middle, or right); and (c) the spatial positioning of the correct match (left, middle, or right). In the case of the all-color condition, one extra element of counterbalancing was included—the spatial positioning of color (i.e., no one particular color predominated in any one particular position). In addition, across one third of the trials, as would be expected by chance, the correct match was the same color as the sample stimulus; on the remaining trials the correct comparison was a different color from the sample.

When subjects had responded correctly on 36 consecutive matching-to-sample training trials, the testing phase began. The color format for experimental stimuli during the testing phase was as follows.

No-color group - Black lettering
All-color group - Color lettering
Color-test group - Color lettering

During this phase of the experiment subjects were tested on the three C-A matching-to-sample tasks. In these tasks, subjects were presented with either C1, C2, or C3 as the sample stimulus, and had to choose from among the three comparison stimuli A1, A2, and A3. Responding in accordance with an equivalence relation was defined as choosing A1 given C1, A2 given C2, and A3 given C3. Subjects received no feedback on their performance. After each response, the screen simply cleared for 2 seconds before the next trial was presented.

Each of the three C-A matching-to-sample tasks was presented 12 times in one quasi-randomly ordered block of 36 trials. The predetermined quasi-random order of presentation (see Appendix 2) was the same for every subject. Each of the following elements was counterbalanced: (a) the order of presentation of the three MTS tasks; (b) the spatial positioning of the comparison stimuli (left, middle, or right); and (c) the spatial positioning of the correct stimulus in terms of the experimenter-designated equivalence relation (left, middle, or right). In the case of the all-color and color-test groups, which presented color stimuli during testing, one extra element of counterbalancing was included—the spatial positioning of color. In addition, in both conditions the correct stimulus choice, in terms of the equivalence relation, was never the same color as the sample stimulus.

After the subject had completed the 36 trials of the testing phase, the following message appeared on the monitor screen:

> That is the end of the experiment.
> Please contact the experimenter.

If this was the subject's first experimental session, then she or he was reexposed to a second session, during which the subject was reexposed to exactly the same training and testing procedures. If this was the subject's second experimental session, then she or he was thanked and debriefed.

Results

Training

The mean numbers of training trials required during the first exposure were 347.4 (*SD* = 102.2) for the no-color group, 499.1 (*SD* = 532.5) for the all-color group, and 443.8 (*SD* = 344.5) for the color-test group. The corresponding figures for the second exposure were 86.1 (*SD* = 41.0), 60.2 (*SD* = 31.2), and 85.8 (*SD* = 62.9) respectively. A 2-factor analysis of variance revealed a highly significant difference between Exposures 1 and 2, $F[1, 42] = 21.36$, $p = 0.0001$. However, there were no significant differences found between the groups, and no significant interaction effects emerged between groups and exposures.

Testing

The test data were first analyzed in terms of the number of responses emitted by each subject that were in accordance with the designated equivalence relations (defined hereafter as correct responses). Across both test exposures, correct responses on the equivalence tests in the no-color and all-color groups ranged between 3 - 36 and 8 - 36 respectively, with 7 subjects from among these groups scoring more than 30 (83.3%) on at least one exposure (6 subjects, 3 from each group, passed at 90%). In contrast, the range in the color-test group was 0 - 15 (i.e., the maximum score was 41.3%) and, indeed, 3 subjects in this latter group failed to score in either exposure while a 4th scored a total of just 1.

The mean numbers of correct (equivalence) responses in the first exposure were 16 (*SD* = 13.8) for the no-color group, 19.6 (*SD* = 9.8) for the all-color group, and 4.9 (*SD* = 6) for the color-test group (see Figure 1). The corresponding figures for the second exposure were 17.4 (*SD* = 13.8) for the no-color group, 18.6 (*SD* = 15.3) for the all-color group, and 4.9 (*SD* = 5.7) for the color-test group (see Figure 1). A 2-way analysis of variance with the three color groups as Factor A and the two test exposures as Factor B showed a significant effect for groups, $F[2, 42] = 6.51$, $p = 0.003$, but not for exposures. No significant interaction effect was identified. Post hoc analyses (Fisher PLSD) revealed significant differences (a) between the no-color and color-test groups ($p = 0.005$) and (b) between the all-color and color-test groups ($p = 0.0002$).

To determine whether the number of training trials predicted performance on the equivalence tests, two Kendall Rank correlation coefficient tests were conducted (one per exposure) using numbers of training trials required and numbers of correct (equivalence) responses as the X and Y variables. For both exposures, there was a small, nonsignificant negative correlation between the two variables ($r = -0.129$, $p = 0.375$ and $r = -0.165$, $p = 0.257$, for Exposures 1 and 2, respectively).

Another important area for analysis was the extent of nonarbitrary relational responding by subjects in the all-color and color-test groups (see Figure 2). If, as suggested by RFT, nonarbitrary sameness may act as a particularly powerful stimulus dimension in matching-to-sample

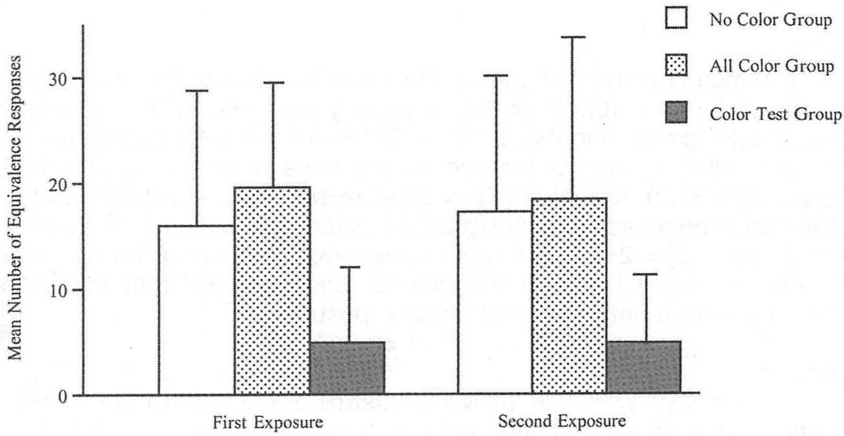

Figure 1. Means and standard deviations for number of responses in accordance with equivalence per individual exposure across the three groups.

contexts, then levels of color matching should have been much higher among subjects in the color-test group than among subjects in the all-color group. For the first exposure, the mean numbers of responses in accordance with the nonarbitrary relation of color were 8.8 ($SD = 6.1$) for the all-color group and 24.8 ($SD = 11.0$) for the color-test group. The corresponding means for the second exposure were 8.9 ($SD = 7.5$) for the all-color group and 25.6 ($SD = 11.0$) for the color-test group. A two-way analysis of variance with the color groups as Factor A and the test exposures as Factor B showed a highly significant effect for groups, F(1,

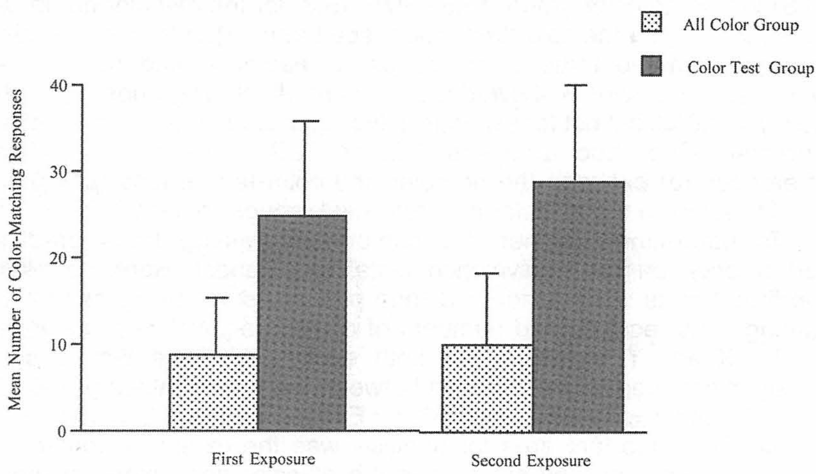

Figure 2. Means and standard deviations for number of responses in accordance with the non-arbitrary relation of color per individual exposure for both the all-color and color-test groups.

28) = 21.99, $p = 0.0001$. There was no significant difference between exposures (Factor B), nor was there any interaction effect. In other words, subjects in the all-color group, for whom reinforcement during training was contingent on ignoring color, showed a significantly lower rate of nonarbitrary relational responding than those in the color-test group (who were not trained to ignore color).

One final feature of the data we examined in greater detail was the extent to which subjects who failed the equivalence tests responded to the nonarbitrary properties of the nonsense syllables (other than color). For example, we determined whether subjects picked A1 (ZID) in the presence of C2 (BIF) based on the common vowel that occurs in these nonsense syllables. No consistent response patterns based on these extra-chromatic, nonarbitrary properties emerged across subjects.

Discussion

The main findings were as follows. Significantly higher levels of equivalence responding were observed for the no-color and all-color groups relative to the color-test group. Furthermore, there was a significantly higher level of nonarbitrary relational responding or color matching in the color-test group than in the all-color group. These results are consistent with the RFT prediction, that nonarbitrary relational responding, in certain contexts, may interfere with arbitrary relational responding. Of course, it may be possible to predict at least some of these results using theoretical approaches other than RFT. We have not done so here, however, because the current work was produced directly by Relational Frame Theory which forms the core of our ongoing research program.

One possible criticism of the present research is that the markedly different mean performance level demonstrated by the color-test group might simply have been the result of the previously unencountered physical property of color in the experiment. If this had been the case, however, then the response errors made by this group would have tended to be of a random nature. Instead, the experimental results show a pattern in which there is a bias towards responding in accordance with the nonarbitrary relation of color. Such an outcome would be highly unlikely on the basis of chance alone (i.e., for both exposures, the mean number of color-matching responses was 24+, well above the 12 responses that would be expected due to chance).

One might argue that the effects reported in this study would disappear entirely with procedures that produced robust emergence of untrained relations. Perhaps, for example, at least some of the subjects in the color-test group would have passed the equivalence test if they had been exposed to additional training and testing. Notwithstanding this possibility, it is important to understand that the purpose of the current work was to demonstrate interference with the *emergence* of derived relational responding by nonarbitrary stimulus relations, rather than the disruption of already well established equivalence relations. As such, the present study constitutes a successful attempt to isolate a variable that may compete with the emergence of equivalence responding, and thus these findings supplement the relatively limited data available on the disruption of equivalence class formation.

Although the current data are relatively clear, it is important to recognize that the current study adopted only one of many different training protocols that have been reported in the literature on derived stimulus relations. For example, some researchers have reported that testing for symmetry before testing for transitivity and combined symmetry and transitivity may facilitate the emergence of equivalence relations (Adams, Fields, & Verhave, 1993). Symmetry relations were not tested in the current study and thus it remains to be seen what effect if any such tests might have on the group effects reported here. Similarly, various training and testing designs have been found to differentially affect testing outcomes. For instance, there are grounds on which to suspect that the so-called linear training design (i.e., train A to B, and B to C), adopted in the current study, may not be the most effective method for establishing equivalence relations (see Arntzen & Holth, 1997, for a detailed discussion of this issue). Once again, future research might determine if the differences observed in the current study are replicated with different designs. Similar questions might also be asked in relation to the use of different types of instructions, and perhaps different subject populations. One of the difficulties in approaching these issues in a systematic manner, however, is that the effect of the foregoing variables on equivalence class formation has yet to be completely and clearly documented. In any case, the very clear effects reported in the current study indicate that the role of nonarbitrary relational responding is an important area for future research.

The current study may provide us with a possibly useful means of examining motivational variables in the context of derived stimulus relations. More specifically, it could be argued that the color-test group were presented with tests in which two consistent patterns of responding could emerge; one based on color matching and the other on equivalence relations. According to RFT, the latter performance requires greater behavioral 'effort' than the former (due to the likely different behavioral histories involved). Insofar as this is the case, an interesting question concerns how one might encourage subjects to engage in 'greater behavioral effort' (i.e., demonstrate equivalence rather than color matching when both patterns of responding are possible). Clearly, historical variables would play an important role here. For example, imagine that the subjects were first successfully trained and tested for equivalence responding using 'standard' no-color stimuli. Would the behavioral momentum created by a history of equivalence responding override the tendency seen in the current study to produce the least effortful test performance? The concept of behavioral momentum suggests that this would in fact occur.

In summary, the current research has shown that the juxtaposition of conflicting arbitrary and nonarbitrary relations may result in significant interference with arbitrary equivalence responding. In a more general sense, what these results indicate is that there are certain conditions under which the interplay between arbitrary and nonarbitrary relational response domains may be significant. In looking at this particular type of interplay, this research is the first of its kind. Given the relative clarity of the current data, and their support for RFT, the approach taken here may well be of use to other researchers concerned with the behavior-analytic investigation of human language and cognition.

References

ADAMS, B. J., FIELDS, L., & VERHAVE, T. (1993). Effects of test order on intersubject variability during equivalence class formation. *The Psychological Record*, 43, 133-152.

ARNTZEN, E., & HOLTH, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record*, 47, 309-320.

BARNES, D., BROWNE, M., SMEETS, P. M., & ROCHE, B. (1995). A transfer of functions and a conditional transfer of functions through equivalence relations in three- to six-year-old children. *The Psychological Record*, 45, 405-430.

BARNES, D., & HOLMES, Y. (1991). Radical behaviorism, stimulus equivalence, and human cognition. *The Psychological Record*, 41, 19-31.

BARNES, D., & KEENAN, M. (1993). A transfer of functions through derived arbitrary and non-arbitrary stimulus relations. *Journal of the Experimental Analysis of Behavior*, 59, 61-81.

BARNES, D., LAWLOR, H., SMEETS, P. M., & ROCHE, B. (1995). Stimulus equivalence and academic self-concept among mildly mentally handicapped and nonhandicapped children. *The Psychological Record*, 46, 87-107.

BARNES, D., & ROCHE, B. (1996). Relational frame theory and stimulus equivalence are fundamentally different: A reply to Saunders' commentary. *The Psychological Record*, 46, 489-507.

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 Behavio*r, 53, 345-358.

FIELDS, L., REEVE, K. F., ADAMS, B. J., & VERHAVE, T. (1991). Stimulus generalization and equivalence classes: A model for natural categories. *Journal of the Experimental Analysis of Behavior*, 55(3), 305-312.

HAYES, S. C. (1991). A relational control theory of stimulus equivalence. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 19-40). Reno, NV: Context Press.

HAYES, S. C., BARNES-HOLMES, D., & ROCHE, B. (2001). Relational Frame Theory: A precis. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian analysis of human language and cognition*. Reno, NV: Context Press.

SIDMAN, M., (1992). Equivalence relations: Some basic consideration. In Steven C. Hayes & Linda J. Hayes (Eds.), *Understanding verbal relations* (pp. 15-27). Reno, NV: Context Press.

WATT, A., KEENAN, M., BARNES, D., & CAIRNS, E. (1991). Social categorization and stimulus equivalence. *The Psychological Record*, 41, 33-50.

Appendix 1: Training Trials

1 A2(R)*	2 A1(B)	3 A3(G)	4 A1(G)
B2(R) B3(B) B1(G)	B2(B) B1(R) B3(G)	B1(B) B3(R) B2(G)	B3(B) B2(R) B1(G)
5 A3(R)	6 A2(B)	7 A3(B)	8 A2(G)
B3(G) B1(R) B2(B)	B1(B) B2(G) B3(R)	B1(G) B2(R) B3(B)	B3(G) B2(B) B1(R)
9 A1(R)	10 A3(R)	11 A1(G)	12 A2(G)
B1(R) B3(B) B2(G)	B2(R) B3(B) B1(G)	B3(G) B1(R) B2(B)	B1(B) B3(R) B2(G)
13 A3(B)	14 A1(R)	15 A2(B)	16 A1(B)
B2(B) B1(G) B3(R)	B1(B) B2(R) B3(G)	B3(R) B1(B) B2(G)	B2(G) B3(B) B1(R)
17 A2(R)	18 A3(G)	19 B1(R)	20 B2(B)
B2(B) B1(G) B3(R)	B3(G) B2(B) B1(R)	C3(B) C1(G) C2(R)	C3(R) C2(G) C1(B)
21 B3(G)	22 B2(R)	23 B1(G)	24 B3(B)
C3(R) C1(B) C2(G)	C1(G) C2(R) C3(B)	C1(B) C3(G) C2(R)	C1(G) C3(B) C2(R)
25 B2(R)	26 B1(R)	27 B3(B)	28 B1(B)
C2(G) C3(R) C1(B)	C2(G) C3(B) C1(R)	C3(G) C2(B) C1(R)	C1(R) C2(G) C3(B)
29 B3(G)	30 B2(G)	31 B1(B)	32 B2(G)
C2(B) C1(R) C3(G)	C3(R) C1(B) C2(G)	C2(B) C1(R) C3(G)	C1(R) C3(G) C2(B)
33 B3(R)	34 B2(B)	35 B3(R)	36 B1(G)
C2(G) C3(B) C1(R)	C2(G) C1(B) C3(R)	C1(G) C2(R) C3(B)	C3(B) C2(R) C1(G)

Appendix 2: Testing Trials

1 C2(G)*	2 C3(B)	3 C2(R)	4 C1(R)
A1(R) A3(G) A2(B)	A3(G) A2(R) A1(B)	A1(G) A2(B) A3(R)	A3(R) A1(B) A2(G)
5 C3(R)	6 C1(B)	7 C2(G)	8 C3(G)
A2(G) A3(B) A1(R)	A1(G) A2(R) A3(B)	A3(B) A1(G) A2(R)	A2(G) A3(R) A1(B)
9 C1(G)	10 C2(B)	11 C1(G)	12 C3(B)
A1(B) A2(G) A3(R)	A1(B) A3(R) A2(G)	A3(R) A2(G) A1(B)	A1(G) A3(R) A2(B)
13 C2(R)	14 C1(B)	15 C3(R)	16 C2(B)
A2(G) A3(B) A1(R)	A2(B) A1(G) A3(R)	A1(R) A2(G) A3(B)	A2(R) A1(G) A3(B)
17 C3(G)	18 C1(R)	19 C2(G)	20 C3(B)
A3(R) A1(B) A2(G)	A2(B) A3(R) A1(G)	A1(R) A3(G) A2(B)	A3(G) A2(R) A1(B)
21 C2(R)	22 C1(R)	23 C3(R)	24 C1(B)
A1(G) A2(B) A3(R)	A3(R) A1(B) A2(G)	A2(G) A3(B) A1(R)	A1(G) A2(R) A3(B)
25 C2(G)	26 C3(G)	27 C1(G)	28 C2(B)
A3(B) A1(G) A2(R)	A2(G) A3(R) A1(B)	A1(B) A2(G) A3(R)	A1(B) A3(R) A2(G)
29 C1(G)	30 C3(B)	31 C2(R)	32 C1(B)
A3(R) A2(G) A1(B)	A1(G) A3(R) A2(B)	A2(G) A3(B) A1(R)	A2(B) A1(G) A3(R)
33 C3(R)	34 C2(B)	35 C3(G)	36 C1(R)
A1(R) A2(G) A3(B)	A2(R) A1(G) A3(B)	A3(R) A1(B) A2(G)	A2(B) A3(R) A1(G)

* The bracketed letters stand for the colors (R = Red, G = Green, B = Blue) in which nonsense syllables appeared. In training trials (Appendix 1), nonsense syllables were colored only for the all-color group, while they were in black lettering for the no-color and color-test groups. In testing trials (Appendix 2) nonsense syllables were colored for the all-color and color-test groups, while they were in black lettering for the no-color group.