Competing Arbitrary and Non-arbitrary Stimulus Relations: The Effect of Exemplar Training in Adult Participants

Neil Kenny[1], Sarah Devlin[2], Dermot Barnes-Holmes[2], Yvonne Barnes-Holmes[2]

&

Ian Stewart[3]

[1]UNIVERSITY OF LIMERICK

[2]NATIONAL UNIVERSITY OF IRLEAND, MAYNOOTH

[3]NATIONAL UNIVERSITY OF IRLEAND, GALWAY

Correspondence relating to this article should be addressed to;
Neil Kenny, Department of Education and Professional Studies,
University of Limerick, Co. Limerick, Republic of Ireland.
Email: Neil.Kenny@ul.ie
Ph: 00353 61 234841

Abstract

The current study sought to extend the work of Stewart et al. (2002) by investigating the effects of differing histories with regard to colour on participant's equivalence test performances. Thirty-two university students were divided into four groups exposed to different training procedures but all were subsequently exposed to an identical equivalence test using coloured stimuli (either red or green). Initially, participants were divided into two groups, one receiving training with stimuli coloured either red or green (the *All Colour* group), while the other received training with stimuli coloured only black (*Colour Test* group). In addition, half of the participants in both groups were exposed to exemplar training procedures (*Exemplar Training Groups)*, with the remaining participants simply receiving repeated test exposures *(Repeat Groups)*. Results showed that participants in the *All Colour* groups produced higher equivalence responding relative to participants in *Colour Test* groups. Exemplar training was shown to be effective in promoting equivalence responding only for participants in the *All Colour g*roups.

The phenomenon of stimulus equivalence was first examined by Sidman (1971) using systematic behaviour analytic methodologies in an attempt to devise improved methods for teaching reading comprehension. His seminal study examined the emergence of untrained, or derived, associations between stimuli using a developmentally disabled participant. The participant was trained to match spoken words to pictures and spoken to printed words before subsequently matching printed words to pictures without additional training. These untrained responses are now often referred to as derived stimulus relations because they emerge in the absence of explicit training. The specific derived relations observed in Sidman's study were labelled equivalence classes, and this label is still used today by most researchers to describe classes of mutually substitutable stimuli (Barnes, Browne, Smeets, & Roche, 1995; Fields, Adams, Verhave, & Newman, 1990; Sidman, 1992).

Behavioural research makes a distinction between responding under the control of arbitrary and non-arbitrary stimulus relations. It has been shown that it is possible for both non-humans and humans to demonstrate control by stimulus relations when such relations are based on the formal properties of the related events. For example, animals may be taught to discriminate relationships such as brighter or dimmer (Reese, 1968), in which the non-arbitrary features of the environment specify the relationship between the stimuli. The phenomenon of stimulus equivalence, and its related effects, indicates that language-able humans also have the ability to show the abstraction of such non-arbitrary relational responding, in that it comes under the control of arbitrary contextual cues available in the environment. In other words, the stimuli involved in the relation are related to each other based on a history of conventional training in derived relations, regardless of the non-arbitrary form of the related events (Hayes, Gifford & Ruckstuhl, 1996). For example, there is nothing in the form of the written word “car” that would allow the formation of a non-arbitrary relation to an actual car. The two stimuli are related together because of a history of reinforcement for derived relational responding, not because the word and the object are physically similar. This type or relational responding has become known as arbitrary relational responding, of which stimulus equivalence is one example.

Relatively early research in the area of stimulus equivalence tested the prediction that the formation of equivalence classes may be disrupted or retarded by introducing competing sources of stimulus control (e.g., Barnes, Lawlor, Smeets, & Roche, 1995; Watt, Keenan, Barnes & Cairns, 1991). In the earliest of these studies participants from Northern Ireland were exposed to a MTS training procedure that would predict the emergence of equivalence relations between Catholic names and Protestant symbols. The results showed, however, that a large proportion of Irish participants chose the untrained Protestant names in the presence of the Protestant symbols. In effect, the pre-experimentally established relation between the Protestant names and symbols interfered with the formation of equivalence relations between the Protestant symbols and Catholic names. In contrast, the results for English participants did not show disruption of equivalence classes because, it was argued, they lacked the pre-experimental exposure to the social contingencies operating in Northern Ireland (Watt, et al., 1991).

This early research focused on competing sources of arbitrary rather than non-arbitrary stimulus control (the latter requires that stimulus relations be defined in terms of the physical properties of the stimuli, rather than purely by social or verbal convention). Only one published study has examined the relationship between non-arbitrary and arbitrary relational responding in a systematic manner (Stewart, Barnes-Holmes, Roche, & Smeets, 2002). In this study, participants were taught three A-B and three B-C matching to sample (MTS) tasks. For the three A-B tasks each participant was presented with A1, A2 or A3 as the sample stimulus and then had to choose among comparisons B1, B2, or B3 (see Figure 1). The correct response was B1 given A1, B2 given A2 and B3 given A3. For the B-C relations the participants received training to match B1-C1, B2-C2, and B3-C3. After this training procedure was completed participants were tested for C-A equivalence responding.

The participants were also divided into three separate groups; No-Colour, All-Colour, and Colour-Test. Each of these groups was exposed to different training and testing procedures. The All-Colour Group was referred to as such because *all* of the stimuli to which the participants were exposed throughout the study were coloured. In this case, the correct comparison stimulus was the same colour as the sample on some trials and a different colour on other trials. In effect, this procedure taught the participant to “ignore” the property of colour as the basis for correct responding. The No-Colour group were exposed to a training procedure identical to the All-Colour group except that black stimuli were used during both training and testing. There was, therefore, no conflicting non-arbitrary relation between the stimuli that could hinder the formation of arbitrary relations. The Colour-Test group were referred to as such because all of the stimuli presented in the basic conditional discrimination training were black, whereas the stimuli presented during the equivalence test were coloured. The objective of this test condition was to determine if participants would respond on the basis of an arbitrary equivalence relation (derived from their *experimental training history*) or on the basis of a non-arbitrary sameness relation of colour when both relations were present. The results of this study indicated, as predicted, that the All-Colour and No-Colour groups produced greater accuracy on the equivalence tests than the Colour-Test group. Furthermore, the errors produced by the Colour-Test group consisted of responses that tended to involve matching the stimuli simply on the basis of colour.

The overall aims of the present study were two-fold. One aim was to extend the work of Stewart et al. (2002) by investigating further the effect of different types of experimental training history, with regard to colour, on participants' performances in equivalence tests. Toward this objective, half of the participants were assigned to an A*ll-Colour* group and, as in the Stewart et al. study, were presented with an MTS training and testing procedure using coloured stimuli throughout. The other half of the participants were assigned to a *Colour-Test* group, similar to that employed by Stewart et al. (i.e., No colour training followed by colour testing). The second key aim of the current study was to determine if equivalence responding with different experimental histories with regard to colour (i.e., All-Colour compared to Colour-Test) could be improved using further training with multiple exemplars.

Past research has provided empirical evidence for the efficacy of exemplar training in establishing repertoires of derived relational responding (Barnes-Holmes, Barnes-Holmes, Roche & Smeets, 2001, Barnes-Holmes, Barnes-Holmes, Smeets, Strand & Friman, 2004). It is possible, therefore, that exemplar training might prove useful in training participants *not* to be influenced by conflicting non-arbitrary relations (Barnes-Holmes, Barnes-Holmes, Roche, & Smeets, 2001). Consider, for example, a participant from the Colour-Test group who persistently fails the equivalence test by responding on the basis of colour. Could subsequent equivalence test performances involving novel coloured stimuli be improved by explicitly training the arbitrary C-A relations across three separate sets of stimuli? To answer this question, half of the participants from both the All-Colour and Colour-Test groups received this type of training. These groups are referred to as Exemplar-Training-All-

Colour (*Exemplar All Colour*) and Exemplar-Training-Colour-Test (*Exemplar Colour Test*), respectively. To determine the efficacy of the exemplar training, the remaining participants from the All-Colour and Colour-Test groups were simply exposed to repeated equivalence test phases – would repeated exposure alone be sufficient to improve equivalence performances? These groups are referred to as Repeated-Test-All-Colour (*Repeat All Colour)* and Repeated-Test-Colour-Test (*Repeat Colour Test*), respectively. The participants were, therefore, divided into 4 groups of equal number. The basic experimental design involving the four groups is presented in Table 1.

Method

Participants

Thirty-two undergraduate students from the National University of Ireland, Maynooth participated in the current study, seventeen were female and fifteen were male. Participants ranged in age from 18-25 years, with a mean age of 20.3 years. All of the participants were recruited through personal contacts from the experimenter and none had any prior knowledge of stimulus equivalence or experimental psychology, and none of them were colour blind. Participants were randomly assigned to one of four experimental groups, with eight participants assigned to each group. None of the participants received any remuneration for taking part in the study and were provided with instructions detailing the general background to the research. Participants were given a general illustration and explanation of the trial types used in the tasks and they subsequently confirmed that they fully understood the instructions by clicking the yes button on the screen.

Setting, Apparatus, and Materials

All participants completed the study in a small quiet experimental room located in the Department of Psychology. Each participant was seated at a desk that contained an iMac computer with a 15-inch colour monitor. Participants interacted with the computer by using three designated keys (Z, V, and M) located on the left, middle, and right of the computer keyboard, respectively. The three response keys were each distinguished with a small yellow sticker. Computer programs controlled all of the stimulus presentations and simultaneously recorded all of the participants' responses. The programs were written in BBC BASIC and were adapted from those employed by Stewart, et al. (2002).

The experimental stimuli were twenty-seven three-letter nonsense syllables that were randomly divided into three sets of nine stimuli, hereafter referred to as Sets 1-3. The nine stimuli contained within each set were then subdivided into three three-member stimulus classes. For ease of communication, each of the twenty-seven nonsense syllables was designated with an alphanumeric label pertaining to a particular class. The organisation of the stimuli and their alphanumeric labels are presented in Table 2. Participants were not exposed to the alphanumeric labels at any time. On some occasions, the nonsense syllables were presented on screen in one of three colours (red, green, or blue), whereas on other occasions they were printed in black only. The presence or absence of stimulus colour was determined by experimental condition (see Table. 1). In addition to the nonsense syllables, participants were provided with a set of instructions typed on an A4 sheet of paper at the beginning of the experiment.

General Procedure

Participants completed the experiment individually, and required only one experimental session that lasted approximately one hour. Once seated at the computer, the participant was verbally instructed as follows:

> Please read the typed instructions on the desk. At various points throughout the experimental session a message will appear on the computer screen. This will signal the end of that part of the experiment. I will be seated outside, please come and inform me when this happens.

The experimenter remained in the room while each participant read the typed instruction sheet containing the following information:

> This is an experiment designed to investigate adult learning and reasoning abilities. There is no "trick" or hidden agenda behind this experiment, it is concerned only with your ability to learn and reason. During the experiment you will be presented with nonsense syllables (VEK, COL, MUJ, etc.). Your task is to learn which one of the three nonsense syllables that appears at the bottom of the screen "goes with" the one at the top.
>
> At first, there is no way you can know which is the correct choice, but each time you choose, the computer will tell you whether you chose the correct or wrong nonsense syllable. With enough experience, you will learn which nonsense syllables go together.
>
> After you have mastered these performances, the computer will then present you with a "test" without feedback (i.e., you will not be told whether your choices were right or wrong). During the test, there is always a correct choice based on what you have learned, so be careful to pay close attention during the first part of the experiment.
>
> Again, it must be stressed that there is no hidden "trick" involved here. Your job is simply to learn, via trial-and-error, which nonsense syllables go together, and then to pass a test based on what the computer has previously taught you.
>
> Please note. You choose the nonsense syllable that appears on the left of the screen 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.

All of the experimental trials in the current study were presented in a MTS format. On each trial, a sample stimulus appeared in the top centre portion of the screen, and three comparison stimuli appeared along the bottom, one on the left, one in the middle, and one on the right. Participants were simply required to choose the comparison that ‘went with’ the sample by pressing the appropriate key on the keyboard. Participants selected the left, middle, or right comparison by pressing the Z, V, or M keys that were located on the left, middle and right of the keyboard, respectively.

The MTS format was used to establish a total of nine three-member equivalence classes, divided across the three stimulus sets – Sets 1-3. Set 1 consisted of the classes A1-B1-C1, A2-B2-C2 and A3-B3-C3; Set 2 consisted of the classes A4-B4-C4, A5-B5-C5 and A6-B6-C6; Set 3 consisted of the classes A7-B7-C7, A8-B8-C8 and A9-B9-C9 (see Table 2). Training within each set required that the participants learn the A-B and B-C conditional discriminations among the three classes. Testing with each set involved presenting each of the three C stimuli as samples with the three A stimuli as comparisons. The training and testing for each set was achieved across three experimental phases. After completing the equivalence test in Phase 3, for each set, participants in Phase 4 received either a second test exposure (i.e., participants in the Test groups) or a second exposure that involved corrective feedback (i.e., participants in the Train groups). Therefore, phase 3 equivalence results are represented in Figure 2 as Exposure 1 for each set and phase 4 equivalence results are represented as Exposure 2 for each of the stimulus sets. In other words, participants were exposed to four experimental phases for each stimulus set. After exposure to Set 1, participants were subsequently exposed to the four phases in Set 2 in the same way and finally to Set 3. A schematic representation of the experimental phases employed for each stimulus set is presented in Figure 1.

Programmed Consequences

A correct response during all training trials was followed by the word "CORRECT", printed in black Arial 72, which appeared in the centre of the screen immediately after the participant's response. This feedback was accompanied by a high-pitched tone. An incorrect response was similarly followed by the word "WRONG", but no explicit tone was audible. Programmed consequences were employed with all training trials, but no programmed consequences followed any test trial.

Procedure

All participants were exposed to the same experimental sequence that consisted of three exposures (one for each stimulus set) to four phases of training and testing. Participants completed Phases 1-4 in that order for each set, before proceeding to the next set.

Phase 1: A-B training. All participants were first exposed to the AB relations in Set 1. This phase consisted of 18 training trials in which the A1-B1, A2-B2, and A3-B3 relations were explicitly trained. Specifically, participants were presented with A1, A2, or A3 as the sample and B1, B2, and B3 as the comparisons. A correct response involved selecting B1 (rather than B2 or B3) in the presence of A1, B2 in the presence of A2, and B3 in the presence of A3.

The block of 18 AB training trials was presented in a pre-determined quasi-randomly ordered sequence that was identical for all participants. What differentiated the participant groups at this point was the presentation of the A and B stimuli in either black lettering only or in colour. For participants in the two Colour-Test groups, all of the A and B stimuli presented during Phase 1 appeared in black only, whereas the stimuli appeared in colour for participants in the two All-Colour groups. As well as counterbalancing the locations of the comparison stimuli, the use of colour required two additional features of counterbalancing to ensure: (1) that the same colour did not consistently appear in a particular location, and (2) that the correct comparison was the same colour as the sample on only one third of the trials. After completing the 18 A-B training trials in Phase 1, participants proceeded immediately to Phase 2, irrespective of their performances on the A-B relations.

Phase 2: B-C training. Training the B-C relations was identical to Phase 1, except that participants were presented with B1, B2, or B3 as the sample and the three C stimuli as comparisons. In this way, the relations B1-C1, B2-C2, and B3-C3 were established for Set 1. Once again, this training consisted of six exposures to each of the three B-C trial-types, and were presented in a pre-determined quasi-randomly ordered sequence that was identical for all participants (the same colour and position counterbalancing was employed as in Phase 1). In order to proceed to Phase 3, participants were required to reach a mastery criterion of 33 correct responses out of the combined 36 A-B and B-C training trials. Participants who did not reach this criterion after the B-C trials were immediately re-exposed to the A-B training followed by the B-C training, and this retraining continued until the mastery criterion was reached. This procedure was employed to ensure that participants had successfully demonstrated the conditional discriminations before proceeding to the equivalence test phase.

Phase 3: C-A testing. The C-A equivalence testing in Phase 3 was identical in format to the previous phases, except that no corrective feedback was provided. The test consisted of 36 trials with twelve exposures to each of three C-A trial-types, in which C1, C2, or C3 was the sample and the three A stimuli were the comparisons. Once again, the CA test trials were presented in a pre-determined quasi-randomly ordered sequence that was identical for all participants. Correct responding in accordance with the designated equivalence relations in Set 1 was defined as selecting A1 in the presence of C1, A2 in the presence of C2, and A3 in the presence of C3 (i.e., the relations C1-A1, C2-A2, and C3-A3 were designated as correct based on the previous training). It is important to emphasise that all participants were presented with the equivalence test trials in colour. No test trials were consequated by feedback of any kind and thus trials were presented one after another with an inter-trial interval of 2 s. Once again, counterbalancing insured that the same colour did not consistently appear in a particular location and the correct comparison was the same colour as the sample on only one third of the trials. All participants proceeded to Phase 4 irrespective of their performances on the equivalence test.

Phase 4: C-A testing or C-A training. The specific nature of Phase 4 to which participants were exposed depended on the group to which they had been assigned. For participants in the *Test* groups (i.e. *Repeat All Colour* and *Repeat Colour Test* groups), Phase 4 was identical to Phase 3 and simply involved a second identical exposure to the equivalence test, using the same stimulus set employed in the previous phase. For participants in the *Train* groups (i.e., *Exemplar All Colour* and *Exemplar Colour Test* groups), Phase 4 was also identical to Phase 3, except that corrective feedback now consequated each trial. In this way, Phase 4 constituted a type of exemplar training (using the stimulus set that had previously been employed in testing). Irrespective of condition and performance, all participants were exposed to the same block of 36 trials in Phase 4 and, thereafter, continued immediately with the rest of the experiment.

After completing Phase 4 with Set 1, all participants were given a short break of two minutes, after which they were immediately exposed to the same training and test sequence with Set 2. All aspects of training and testing in Phases 1-4 with Set 2 were identical to Set 1, except that three novel three-member classes (A4-B4-C4, A5-B5-C5, and A6-B6-C6) were trained and tested. Once again, the use of black or coloured stimuli during the conditional discrimination training in Phases 1 and 2 differentiated the *Colour-Test* and *All-Colour* groups and the use of corrective feedback in Phase 4 differentiated the Train and Test groups. Once participants had completed the C-A trials in Phase 4 with Set 2, they were once again given a two-minute break and thereafter proceeded immediately to training and testing with Set 3.

All aspects of training and testing in Phases 1-4 with Set 3 were identical to the previous sets, except that three novel three-member classes (A7-B7-C7, A8-B8-C8, and A9-B9-C9) were trained and tested. After completing the last block of C-A trials in Set 3, all participants were presented with the following instruction on the computer screen: “That is the end of the experiment. Please contact the Experimenter.” For all participants, this marked the end of the experiment.

Results

All participants in the current study were exposed to a total of six blocks of 36 C-A trials presented across three stimulus sets (two in each set), hereafter referred to as Exposures 1 and 2 in Set 1, Set 2, and Set 3. The mean accuracy scores (with standard error bars) of participants in each of the four groups during Exposures 1 and 2 for each of the three stimulus sets are presented in Figure 2. Notable differences in the equivalence performances of participants in the four experimental groups can be observed. The sixteen participants in the two *All-Colour* groups produced the highest levels of accuracy on equivalence responding. Participants in the *Exemplar All Colour* group produced the most accurate equivalence scores across all three sets with accuracy increasing across sets, particularly after the C-A training in Exposure 1 of Set 1. Mean accuracy scores for this group ranged from 23.12 (*SE* = 5.9) in Exposure 1 of Set 1 to 35.6 (*SE* = 0.74) in Exposure 2 of Set 3. Participants in the *Repeat All Colour* group also produced relatively high levels of accuracy on equivalence, again with an improving trend across sets (mean accuracy scores ranged from 15.12 (*SE* = 8.5) in Exposure 1 of Set 1 to 34.38 (*SE* = 1.6) in Exposure 2 of Set 3.

Participants in the *Exemplar Colour Test* group produced lower levels of accuracy on equivalence overall. On the first test exposure to Set 1, these participants produced only low levels of accuracy but this performance improved greatly in the C-A training provided in the second exposure. These improvements, however, did not appear to be retained in the equivalence test in Set 2, although performances improved again with explicit C-A training. A similar pattern of responding was recorded with these participants on Set 3, with the final performance in Exposure 2 of Set 3 demonstrating a relatively high level of accuracy overall. The mean accuracy scores for this group ranged from 9.62 (*SE* = 11.9) in Exposure 1 of Set 1 to 29.6 (*SE* = 6.0) in Exposure 2 of Set 3, indicating a sizeable improvement. In contrast, participants in the *Repeat Colour Test* group produced the lowest equivalence scores overall and their performances remained relatively low even for Exposure 2 of Set 3. The mean accuracy scores for this group ranged from 9.75 (*SE*=5.5) in Exposure 1 of Set 1 to 19.0 (*SE*=9.1) in Exposure 2 of Set 3.

The results thus far may be summarized as follows. The two All-Colour groups produced more equivalence test responding across all three sets than the Colour-Test groups. Furthermore, the equivalence responding of the All-Colour groups reached near perfect levels by the third set. For the Colour-Test groups, however, equivalence test responding remained relatively low across each set. Interestingly, the explicit C-A training given to the *Exemplar Colour Test* group increased accuracy on the trained set, but it appeared to have little effect on the equivalence responding in the first exposure to the next set. In short, training with coloured stimuli, rather than exemplar training, appeared to a major factor in improving equivalence responding across stimulus sets. Although it should be noted that combining colour and exemplar training produced the most rapid improvement in equivalence responding.

Error Analyses

Colour-matching errors. Inaccurate equivalence responses may reflect two types of responding because each C-A trial contained three comparisons. One comparison was correct in terms of the conditional discrimination training history but was the same colour as the sample on only one third of the trials. Another comparison was the same colour as the sample but was incorrect in terms of the training history on two thirds of the trials. A third comparison was also incorrect but was not the same colour as the sample on all trials (referred to hereafter as the neutral comparison). Figure 3 presents the mean number of error responses for the four groups in which the selected comparison was the same colour as the sample (i.e., non-arbitrary sameness matching). The data show that the colour-matching errors made by the *Repeat All Colour* group clearly decreased across all consecutive exposures to near zero, indicating that repeated test exposures taught participants to ignore colour as a salient feature of the stimuli. The *Exemplar All Colour* group produced fewer colour-matching errors than the *Repeat All Colour* group overall, as would be expected because of the explicit C-A training, and these errors also decreased to near zero by the third set.

The data for both of the Colour-Test groups showed more colour-matching errors when compared to the All-Colour groups. For the *Repeat Colour Test* group, the pattern of colour errors remained high but somewhat variable across the sets, although a general decreasing trend was observed. Participants in the *Exemplar Colour Test* group produced a distinctive pattern of colour-matching responses in which the equivalence tests contained a high number of colour responses, but the explicit C-A training did not. Thus the training decreased colour-matching but this effect was not maintained during subsequent equivalence tests across the three sets.

Neutral errors. Figure 4 presents the participants' errors in which they selected the neutral comparison stimulus. The data show that the neutral errors made by the *Repeat All Colour* group were variable but generally decreasing across sets. The *Exemplar All Colour* group produced fewer neutral errors, a pattern which again decreased consistently across the sets. Both of the Colour-Test groups generally produced more neutral errors than the All-Colour groups and this pattern of neutral errors remained relatively high and variable across the sets.

Discussion

The results of the current study were largely in keeping with the experimental predictions. The participants in the All-Colour groups (who were exposed to coloured stimuli during training and testing) produced superior equivalence performances compared to those participants exposed only to colour stimuli during the C-A trials

(*Repeat Colour Test* group). This finding suggests that a history of training with colour served to undermine inappropriate colour-matching responses during the equivalence tests. The explicit C-A training did improve the equivalence performances of participants who were not trained with coloured stimuli but the improvement failed to transfer in any clear way to subsequent equivalence tests with new stimulus sets. The errors made by the Colour-Test groups were predominantly responses in accordance with non-arbitrary colour-matching, which suggests that the weaker equivalence performances did indeed result from a conflict between the arbitrary and non-arbitrary stimulus dimensions. This finding is entirely consistent with the previous study reported by Stewart et al. (2002).

The design of the current study addresses one argument that may have been forwarded in response to the procedures employed by Stewart et al. (2002). Specifically, the effects produced in the original study were obtained with only one exposure to equivalence training and testing with a single stimulus set, and thus it was not possible to determine if such effects would be maintained across additional training and testing. The current findings clearly indicate that such effects are maintained when participants are not trained with coloured stimuli and even when explicit equivalence training is provided across two exemplars.

An interesting conclusion that emerges from the current study is that exposure to the All Colour conditions appeared to improve C-A equivalence responding to a larger degree than exemplar training *per se.* By Set 3, Exposure 1, both of the All Colour groups were producing near perfect levels of equivalence responding, but the two Colour Test groups were not. In fact, the latter two groups produced very similar performances, thus indicating that the exemplar training had little impact relative to simple repeat exposures. This outcome seems to be at odds with the RFT emphasis on the role of exemplar training in producing derived relational responding (Barnes-Holmes, Barnes-Holmes, Roche & Smeets, 2001). On balance, of course, the experiment involved exposure to three stimulus sets and thus perhaps additional exposures to more exemplars would have produced the predicted effect (see Murphy, Barnes-Holmes, Barnes-Holmes, 2005).

Another strategy for improving equivalence responding in the context of competing non-arbitrary relations might involve exposing participants to training and testing conditions with black stimuli until they demonstrate equivalence responding before introducing a Colour-Test condition. Would competing colour relations undermine previously established equivalence responding within and/or across different sets of stimuli?

Although the current work clearly replicated and extended the study by Stewart et al., (2002), it continues to suffer from some of the weakness highlighted by these researchers in the original study. For example, tests of derived symmetry relations were not included in the current protocol and this may have had a notable impact on the emergence of equivalence responding. Perhaps, for example, equivalence performances of some participants would have been better with the inclusion of symmetry tests prior to the equivalence test. Other researchers have indeed highlighted the importance of the equivalence test sequence in generating strong equivalence performances (Adams, Fields, & Verhave, 1993).

Another important area for possible expansion of the current work would be to determine if these findings could be replicated with different populations. The current study, as with Stewart et al. (2002), was conducted with verbally-sophisticated undergraduate students. Future research could use the procedures employed here with populations such as children diagnosed with autism who may show a high level of interference by non-arbitrary stimulus relations.

As noted in the Introduction, the first study to report interference effects in the formation of equivalence relations involved presenting stimuli that participated in pre-experimentally established social or cultural categories (in Northern Ireland). The current findings do not appear to be directly relevant to this earlier finding because the interference effects reported herein were based on the competing non-arbitrary stimulus relation of color. Nevertheless, the finding that the limited exemplar training employed here appeared to have very little effect on the observed interference effects in the current study may be instructive. That is, if researchers are seeking to employ exemplar training in an effort to override the interference effects of pre-experimentally established social or cultural categories it may be important to employ a relatively large number of relevant exemplars (see Roche, Barnes, & Smeets, 1997).

In summary, the current study has demonstrated that introducing non-arbitrary relations into equivalence test procedures may result in significant interference with equivalence performances. The results also indicate that training with coloured stimuli successfully remediated this interference effect, but the impact of exemplar training with only two stimulus sets had a limited effect on equivalence responding.

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.

Barnes, D., Lawlor, H., Smeets, P. M., & Roche, B. (1995). Stimulus equivalence and academic self-concept among mildly mentally handicapped and non-handicapped children. *The Psychological Record, 46,* 87-107.

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001). Exemplar training and a derived transformation of functions in accordance with symmetry. The Psychological Record, 51, 287-308.

Barnes-Holmes, Y., Barnes-Holmes, D., Smeets, P. M., Strand, P., & Friman, P. (2004). Establishing relational responding in accordance with more-than and less-than as generalized operant behavior in young children. International Journal of Psychology and Psychological Therapy, 4, 531-558.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). Relational frame theory: A précis. In S. C. Hayes, D., Barnes-Holmes, & B. Roche (Eds.), Relational frame theory: A post-Skinnerian account of human language and cognition (pp. 141-154). New York: Plenum.

Hebb, D. O. (1937). The innate organization of visual activity: II. Transfer of response in the discrimination of brightness and size by rats reared in total darkness. *Journal of Comparative Psychology, 24*, 277-299.

Murphy, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2005). Derived manding in children with autism: Synthesizing Skinner’s Verbal Behavior with relational frame theory. Journal of Applied Behavior Analysis, 38, 445-462.

Perkins, F. T. (1931). A further study of configurational learning in the goldfish. *Journal of Experimental Psychology, 14*, 508-538.

Reese, H. W. (1968). The perception of stimulus relations: Discrimination learning and transposition. New York Academic press.

Roche, B., & Barnes, D., & Smeets, P. M. (1997). Incongruous stimulus pairing contingencies and conditional discrimination training: Effects on relational responding. *Journal of the Experimental Analysis of Behavior, 68*, 143-160.

Stewart, I., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2002). Stimulus equivalence and non-arbitrary relations. The Psychological Record, 52, 77-88.

Towe, A. L. (1954). A study of figural equivalence in the pigeon. *Journal of Comparative and Physiological Psychology, 47,* 283-287.

Watt, A., Keenan, M., Barnes, D., & Cairns, E. (1991). Social categorization and stimulus equivalence. The Psychological Record, 41, 33-50.

Appendix 1

TABLES:

Table 1. The Basic Experimental Design Employed in the Experiment.

Participant Groups	**Stimulus Sets**		
	Set 1	*Set 2*	*Set 3*
Repeat-All Colour	CA Test CA Test	CA Test CA Test	CA Test CA Test
Exemplar-All Colour	CA Test CA Train	CA Test CA Train	CA Test CA Train
Repeat-Colour Test	CA Test CA Test	CA Test CA Test	CA Test CA Test
Exemplar-Colour Test	CA Test CA Train	CA Test CA Train	CA Test CA Train

Table 2. The Experimental Stimuli and Designated Alphanumeric Labels Employed in the Experiment.

Stimulus Sets and Classes		
Set 1	*Set 2*	*Set 3*
ZID (A1)	BEM (A4)	WIK (A7)
MAU (B1)	PID (B4)	MUJ (B7)
JOM (C1)	NOS (C4)	JUR (C7)
VEK (A2)	DOK (A5)	LOP (A8)
WUG (B2)	JAD (B5)	ZUN (B8)
BIF (C2)	LUD (C5)	NID (C8)
YIM (A3)	SEP (A6)	SOM (A9)
DAX (B3)	COL (B6)	CAJ (B9)
PUK (C3)	ZUR (C6)	DUP (C9)

Appendix 2

FIGURES:

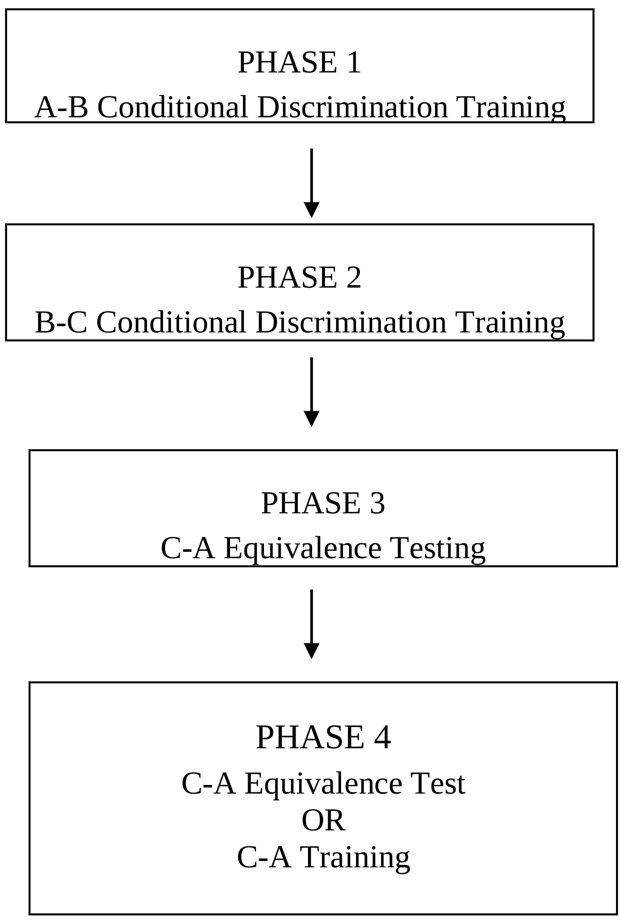

Figure 1. A schematic representation of the experimental phases conducted within each stimulus set.

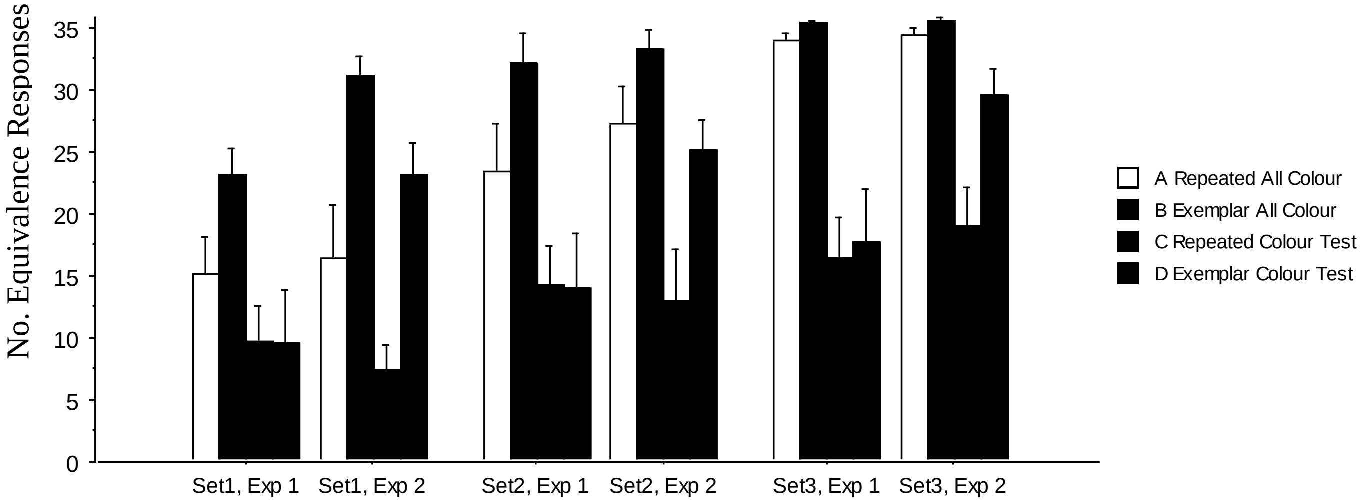

Figure. 2. Equivalence-consistent responding for all groups across three stimulus sets.

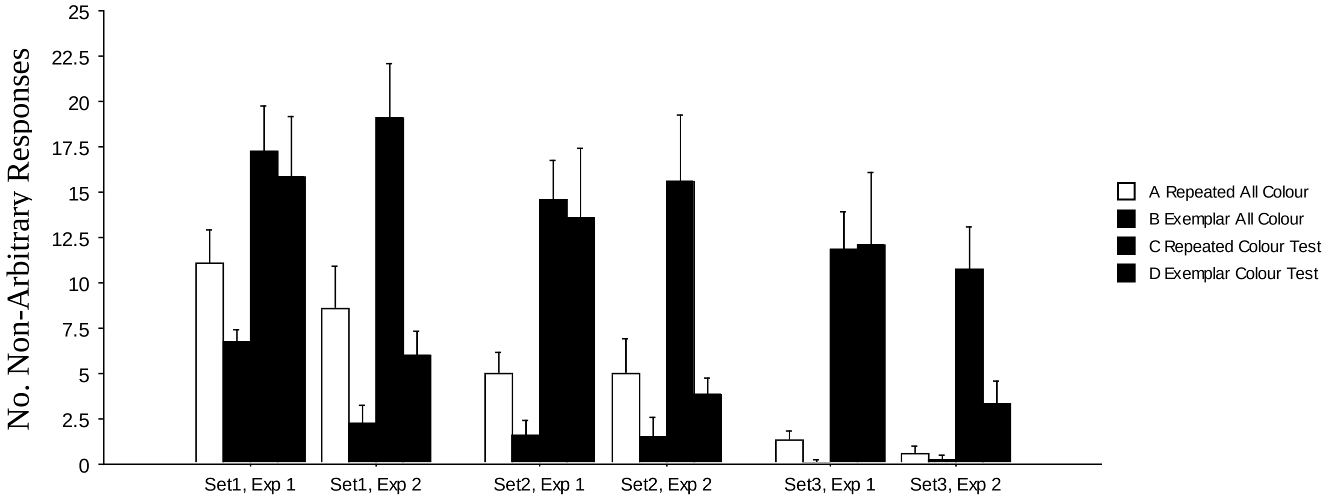

Figure. 3. Number of responses according to non-arbitrary relation by all groups across three stimulus sets

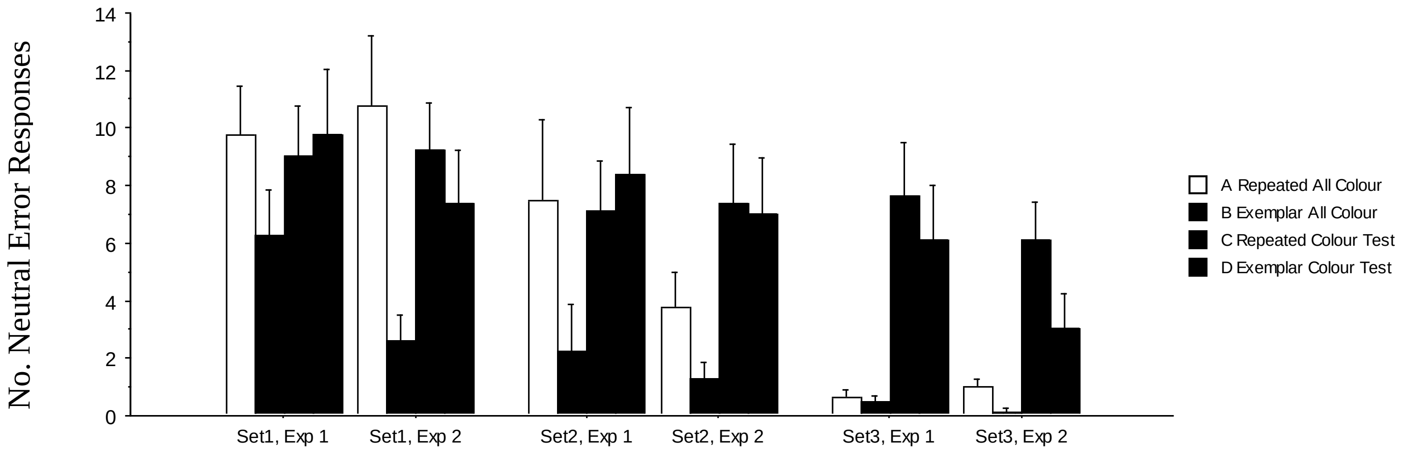

Figure. 4. Number of neutral errors for all groups across three stimulus sets.