

BRIEF REPORT**BUILDING CONTEXTUALLY-CONTROLLED EQUIVALENCE CLASSES TO TEACH ABOUT INFERENTIAL STATISTICS: A PRELIMINARY DEMONSTRATION**

Daniel M. Fienup, Thomas S. Critchfield, Daniel P. Covey

ILLINOIS STATE UNIVERSITY

Ever since a seminal report by Sidman and Cresson (1971), principles of stimulus equivalence have been used to establish skills of practical benefit. As in that seminal report, most applications have focused on the needs of individuals with disabilities (e.g., Stroman, Mackay, & Stoddard, 1992). The present report describes the first steps in an attempt to employ principles of stimulus equivalence to academic instruction in higher education. We will describe how the intervention contributes to instructional efficiency by harnessing the capacity of stimulus equivalence to generate emergent (untaught) abilities, and, in keeping with the notion that translation informs both basic and applied science (Mace, 1994), also derive some useful thoughts about a conceptual issue concerning contextual control (e.g., Bush, Sidman, & de Rose, 1989) over equivalence class membership.

The present application taught selected concepts related to statistical inference and hypothesis decision-making, which some observers regard as among the most difficult topics of instruction in the undergraduate psychology curriculum (e.g., Knowles, 1974). Although students completed all parts of their participation in one "classroom" visit, here, for clarity of exposition, we describe different parts of the intervention as separate studies. Experiment 1 involved the formation of two equivalence classes that were inspired by academic concepts. As we will explain, however, one of these concepts, though structurally sound within the equivalence-class

framework, was academically incomplete. Its shortcomings were the focus of Experiment 2, in which contextual control over stimulus-class membership was established – a novel academic application as far as we are aware. Together these studies describe the preliminary shaping of a repertoire relevant to statistical inference. Finally, in the General Discussion, we discuss some practical and conceptual issues that derive from this translational effort.

GENERAL METHOD

Participants and Setting

Twelve college undergraduate students volunteered, provided informed consent, and scored no better than 67% correct on pretests. The two studies required between about 1 and 1.5 hr of student time, for which they received course bonus credit. The study took place in a classroom containing IBM-compatible computer work stations, using lessons that were created using Visual Basic® 2005 (Dixon & MacLin, 2003) and ran on the Microsoft Windows XP® operating system. Details of programming that, for economy of presentation, are not described here are available from the authors upon request.

General Structure of the Lessons

Learning stimuli and match-to-sample trials. Table 1 shows the stimuli used in the study and the notation assigned to them for ease of exposition here (students were never exposed to this notation). Stimuli within a set were associated with each other during the study. Lesson 1 of Experiment 1 used the A, B, and C stimuli. Lesson 2 of Experiment 1 used the D, E, and F stimuli. In Experiment 2, students were required to attend to A stimuli to make decisions about how D stimuli were related to the E and F stimuli. Note that for some students there were three separate versions of the

Experiment 2 was part of a doctoral dissertation conducted at Illinois State University by Dan Fienup, who is now affiliated with the May Institute.

Correspondence to: T. Critchfield, Psychology Department, Campus Box 4620, Illinois State University, Normal, IL 61790. Electronic mail: tscritic@ilstu.edu.

Table 1
Stimuli.

Notation	Set 1	Set 2
A	Low p -value	High p -value
B	Statistically significant	Not statistically significant
C	$p \leq .05$	$p > .05$
Dx	Scientific Hypothesis: The IV will increase the DV Results: The DV increased	Scientific Hypothesis: The IV will increase the DV Results: The DV did not increase
Dy	Scientific Hypothesis: The IV will decrease the DV Results: The DV decreased	Scientific Hypothesis: The IV will decrease the DV Results: The DV did not decrease
Dz	Scientific Hypothesis: The IV will change the DV Results: The DV changed	Scientific Hypothesis: The IV will change the DV Results: The DV did not change
E	Consistent with Scientific Hypothesis	Not consistent with Scientific Hypothesis
F	Reject Null Hypothesis	Fail to Reject Null Hypothesis

D stimuli, representing different types of predictions about changes in a dependent variable.

The stimuli were displayed in black font within white boxes, and were based on concepts of inferential statistics and hypothesis testing (Huck, 2000). Figure 1 shows how the stimulus boxes appeared during the match-to-sample trials that comprised the lessons: simultaneous presentations of a sample stimulus (top box) and comparison stimuli (lower boxes). On each trial one comparison stimulus showed the correct choice, one showed an incorrect choice, and the third stimulus was a blank (white) box that,

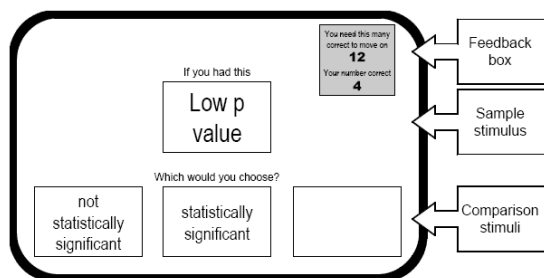


Figure 1

if selected, also counted as an incorrect choice. Table 2 shows the stimuli that were used in each of the 20 learning units across the two studies (notation is as defined in Table 1). Within each learning unit steps were taken to assure unpredictable variation across trials of the screen location of correct comparison stimuli and the stimulus class from which sample and comparison stimuli were drawn. Selection of a comparison stimulus was followed immediately (no intertrial interval) by the next scheduled event, either feedback (see below) or the next trial.

Mastery criteria. Each lesson included a training phase that was organized into two or more blocks of trials, or learning units. Mastery was defined as making correct responses on 12 consecutive trials; thus, each student completed each learning unit scoring at 100% accurate. To complete a lesson, a student also had to demonstrate mastery (89% correct or higher) on a posttest on which no feedback was provided. A student who met this criterion proceeded to the next scheduled task. The protocol required a student who scored lower to repeat the training phase and attempt the posttest again; however,

Table 2
Types of trials used in the learning units.

Experiment	Lesson	Unit	First Trial Type		Second Trial Type	
			Sample	Correct Comparison	Sample	Correct Comparison
1	1	1	A1	B1	A2	B2
		2	C1	A1	C2	A2
	2	1	D _x 1	E1	D _x 2	E2
		2	D _x 1	F1	D _x 2	F2
		3*	D _y 1	E1	D _y 2	E2
		4*	D _y 1	F1	D _y 2	F2
		5*	D _z 1	E1	D _z 2	E2
		6*	D _z 1	F1	D _z 2	F2
2	3	1	A1 + D _x 1	E1	A1 + D _x 2	E2
		2	A2 + D _x 1	E2	A2 + D _x 2	E2
		3	A1 + D _x 1	F1	A1 + D _x 2	F2
		4	A2 + D _x 1	F2	A2 + D _x 2	F2
		5	A1 + D _y 1	E1	A1 + D _y 2	E2
		6	A2 + D _y 1	E2	A2 + D _y 2	E2
		7	A1 + D _y 1	F1	A1 + D _y 2	F2
		8	A2 + D _y 1	F2	A2 + D _y 2	F2
		9	A1 + D _z 1	E1	A1 + D _z 2	E2
		10	A2 + D _z 1	E2	A2 + D _z 2	E2
		11	A1 + D _z 1	F1	A1 + D _z 2	F2
		12	A2 + D _z 1	F2	A2 + D _z 2	F2

Note. In Lesson2 of Experiment 1, learning units designated with an asterisk (*) were completed only by Cohort 2. See text for details.

in this study all students passed the posttests on the first attempt.

Feedback. During learning units, selecting a comparison stimulus immediately produced auditory accuracy feedback (an ascending sound for correct responses and a descending sound for incorrect responses) through stereo headphones, followed by the next trial. During testing phases, clicking on any comparison

stimulus immediately initiated the next trial (no feedback). During both learning units, a blue box located in the upper right corner of the screen (Figure 1) showed the mastery criterion and the number of consecutive correct responses. During tests, the box was red and showed both the number of trials on the test and the number of trials that the student had completed.

EXPERIMENT 1: BASIC EQUIVALENCE CLASSES

The vocabulary of inferential statistics contains several expressions of roughly interchangeable meaning, including that a *low p-value* usually is defined as $p \leq .05$, which indicates that a research result is *statistically significant*. Thus, mastering this subject matter requires grasping the equivalence of these ideas. Similarly, certain features of hypotheses and research results, considered jointly, promote decisions to reject null hypotheses and embrace scientific hypotheses. In this study, students learned conditional discriminations that were expected to promote the formation of limited equivalence classes involving concepts of inferential statistics and hypothesis decisions, including emergent relations abilities that would enhance instructional efficiency. To be clear, there is nothing novel in the structure or genesis of these classes other than the use of specific academically-relevant stimuli. Experiment 1 demonstrated the feasibility of creating these classes, which were a prerequisite to the contextual training of Experiment 2.

METHOD

Overview

The students were run in two cohorts of six students each, both of which completed two lessons, each consisting of a training phase and a posttest, as described below. Both cohorts completed Lesson 1, which was based on concepts of inferential statistics (in Table 1, the A-B-C stimuli). The two cohorts completed similar, but not identical, versions of Lesson 2, which was based on hypotheses decision making (in Table 1, the D-E-F stimuli).

Lesson 1: A-B-C Relations (Statistical Significance)

In two learning units (Table 2), students learned how the following ideas relate: *p-value* descriptors (A stimuli), statistical significance/non-significance (B stimuli), and specific ranges of *p-values* (C stimuli). These are foundational labels underpinning inferential-statistics concepts (Huck, 2000). In Figure 2 (top), directly taught relations are shown as solid arrows and expected emergent transitive relations are shown as gray arrows. In the latter case, on the posttest, students were expected to

demonstrate untaught relations between the B stimuli (*statistically significant* and *not statistically significant*) and the C stimuli ($p \leq .05$ to $p > .05$, respectively).

Lesson 2: D-E-F Relations (Hypothesis Decisions in the Absence of Statistical Information)

Students learned how the following ideas relate: scientific hypothesis paired with a qualitative description of effects (D stimuli), decisions regarding the scientific hypothesis (E stimuli), and decisions regarding the null hypothesis (F stimuli). These are foundational ideas underpinning hypothesis decision making (Huck, 2000). Figure 3 shows the directly taught relations of Lesson 2 as black arrows and expected emergent transitive relations as gray arrows. For economy of presentation, the D stimuli are not shown exactly as students saw them; see Table 1 for the exact stimuli.

During D→E training, students were taught to match comparison stimuli *consistent with the scientific hypothesis* and *not consistent with the scientific hypothesis* with samples in which an effect did or did not, respectively, qualitatively match the prediction of the scientific hypothesis. A note about this phrasing is in order. Although statisticians typically refer to rejecting or failing to reject the scientific hypotheses (Huck, 2000), the present E stimuli were structured for consistency with instruction in an introductory psychological statistics course at the university at which the study was conducted. Thus, where the scientific hypothesis was concerned, *not consistent with the scientific hypothesis* replaced the more traditional *reject*, and *consistent with the scientific hypothesis* replaced *fail to reject*. Presumably, different language would be employed in teaching at different institutions.

During D→F training, students were taught to match comparison stimuli *reject the null hypothesis* and *fail to reject the null hypothesis* with samples in which an effect did or did not, respectively, qualitatively match the prediction of the scientific hypothesis.

On the Lesson 2 posttest, students were expected to demonstrate untaught relations between the E stimuli (*consistent with the scientific hypothesis* and *not consistent with the scientific hypothesis*) and the F stimuli (*reject the*

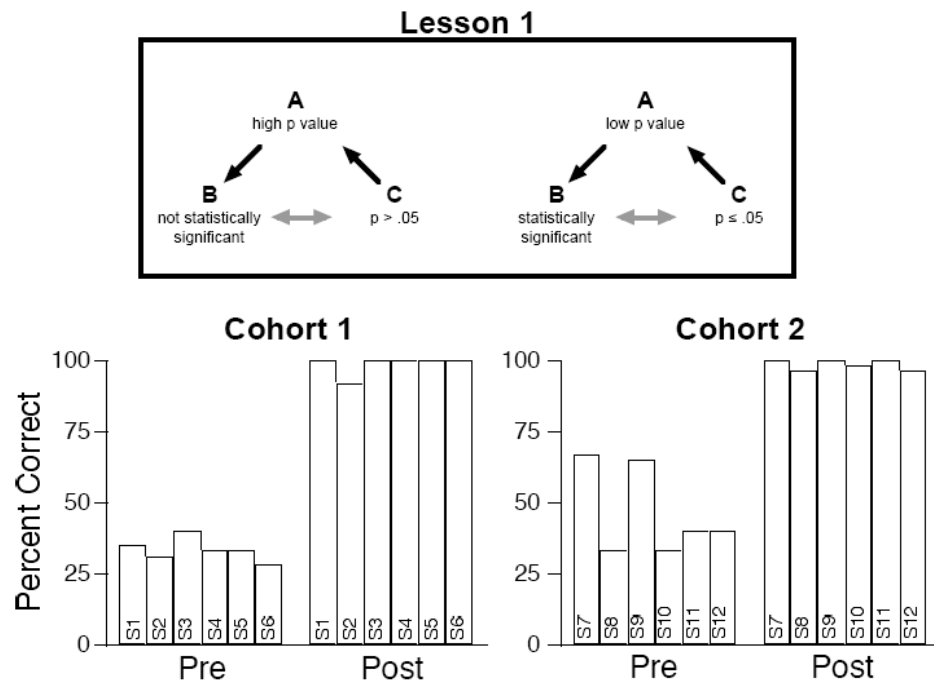


Figure 2

null hypothesis and *fail to reject the null hypothesis*, respectively).

The two cohorts of students differed in terms of the complexity of the D stimuli to which they were exposed. In general, the purpose of the D stimuli was to present situations in which the direction of an effect either did or did not correspond to the scientific hypothesis prediction. Predictions and effects may represent several kinds of effects in a dependent variable (increase, decrease, and change), but for Cohort 1, the scientific hypothesis of the D stimuli always predicted that the dependent variable would *increase*. The corresponding results indicated that the dependent variable *increased* or *did not increase* (the D_x stimuli in Table 1). Thus, students in Cohort 1 completed two learning units (one each for $D \rightarrow E$ and $D \rightarrow F$). For Cohort 2, there were three versions of each D stimuli (the D_x , D_y and D_z stimuli in Tables 1 and 2), with the hypothesis predicting that a dependent variable would *increase*, *decrease*, or *change*, respectively. The corresponding result indicated that the dependent variable did, or did not, show the predicted effect. Thus, students in Cohort 2 completed 6 learning units (for $D \rightarrow E$ and $D \rightarrow F$,

one learning unit for each version of the D stimulus).

RESULTS AND DISCUSSION

Figures 2 and 3 (bottom panels) summarize the results of the pretests and posttests for the two lessons. As noted previously, accuracy was low on the pretests. Subsequently, the students readily mastered the learning phases of the two lessons, typically requiring fewer than 20 trials to meet the mastery criterion for a given learning unit. On posttests, which included both the relations that were directly taught and several expected emergent relations (Figures 2 and 3, top), all students scored near 100%. Thus, as expected in stimulus equivalence, the direct teaching of four relations ($A \rightarrow B$, $C \rightarrow A$, $D \rightarrow E$ and $D \rightarrow F$) promoted the mastery of many others. Traditionally, researchers have counted among the emergent relations the symmetrical, or inverse, variant of the trained relations (here, $B \rightarrow A$, $A \rightarrow C$, $E \rightarrow D$ and $F \rightarrow D$) as well as transitive relations, or those between stimuli that have not been experienced together (here, $B \rightarrow C$, $C \rightarrow B$, $E \rightarrow F$ and $F \rightarrow E$). By this accounting, teaching four relations spawned the mastery of 12. Thus, this study demonstrates the

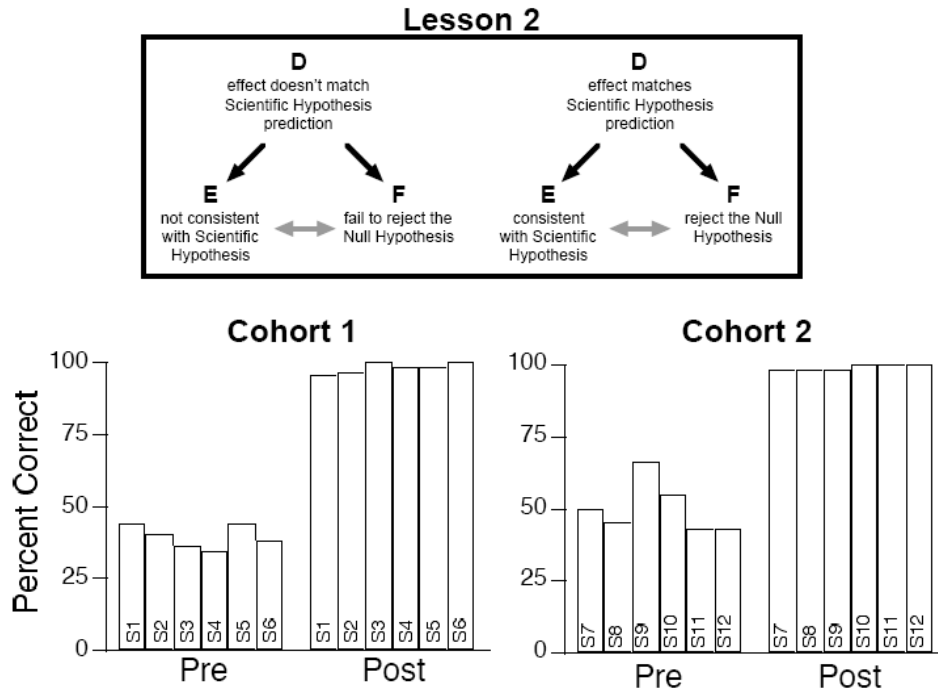


Figure 3

instructional efficiency that is expected of teaching based on stimulus equivalence. Unlike most previous reports, it does so for advanced learners of relatively sophisticated subjected matter.

EXPERIMENT 2: ESTABLISHING CONTEXTUAL CONTROL

An observer of statistical inference will note that relations of Experiment 1 (Lesson 2) inaccurately portray the hypothesis decision process. That lesson taught relations that were unconditional; that is, students were taught to evaluate hypotheses based only on the correspondence between scientific hypothesis prediction and direction of effects. By contrast, relations involved in hypothesis decisions are conditional. They are based on *both* the correspondence between scientific-hypothesis prediction and direction of effects *and* inferential- statistics information (Huck, 2000).

Correcting for this shortcoming was the focus of an additional lesson, which only the students in Cohort 2 completed, and which was designed to promote conditional consideration of both hypothesis information and statistical information. Basic researchers refer to "conditional reasoning" as *contextual control*

(Sidman, 1994), and laboratory studies show that it can be instated via procedures of stimulus equivalence (e.g., Bush, et al., 1989) and suggest that contextual principles derived from abstract stimuli also apply to learning about stimuli of everyday relevance (e.g., Kohlenberg, Hayes, & Hayes, 1991).

In lay terms, contextual cues may be said to "switch on" and "switch off" membership of a particular stimulus in multiple stimulus classes. In the present case, the stimuli of interest were those in the D-E-F class that was established in Experiment 1 (Lesson 2). As Figure 3 (top) shows, in Experiment 1 (Lesson 2) students learned that an effect that matches the scientific-hypothesis prediction (D stimulus) is both *consistent with the scientific hypothesis* (E stimulus) and the occasion on which to *reject the null hypothesis* (F stimulus). In Experiment 2 the students learned that this class is valid only when the D stimulus is accompanied by a *low p-value* (A stimulus). When accompanied by a *high p-value* (A stimulus), an effect that matches the scientific-hypothesis prediction (D stimulus) instead is *inconsistent with the scientific hypothesis* (E stimulus) and the occasion on

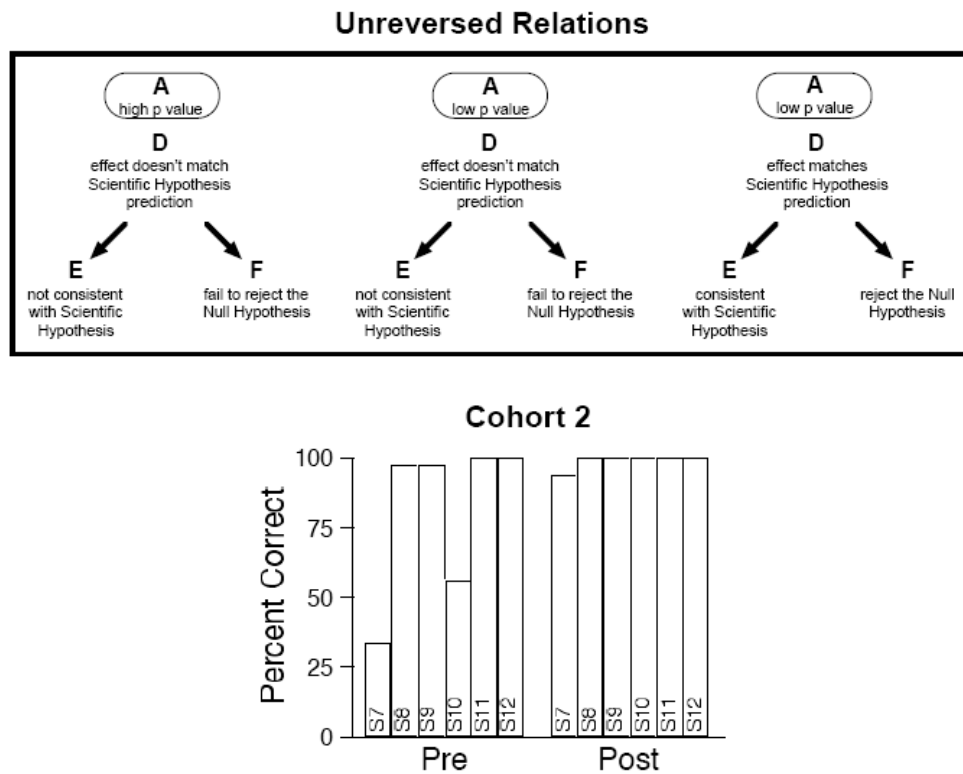


Figure 4

which to *fail to reject the null hypothesis* (F stimulus).

METHOD

General procedures were identical to those of Experiment 1 (Cohort 2) except as noted here. The lesson employed compound sample stimuli in which the D stimuli of Lesson 2 (scientific hypothesis plus a qualitative description of the direction of an effect) were presented simultaneously with A stimuli of Lesson 1 (*low p-value* and *high p-value*). Students were instructed to “use both pieces of information to make decisions.” Figures 4 and 5 (top panels) provide examples; see Table 2 for a full accounting of relation types. There were 12 learning units, reflecting all possible combinations of three types of D stimuli (Table 1), the A stimuli (*low p-value* or *high p-value*); and possible comparison stimuli (E stimuli or F stimuli). Unlike in Experiment 1, the purpose of the pretest and posttest was simply to verify the acquisition of relations that were explicitly taught; no emergent relations were tested. As will be discussed shortly, however, on the basis

of Lesson 3 training a number of additional emergent relations are expected that we did not evaluate.

RESULTS AND DISCUSSION

Students readily mastered the learning units, typically requiring fewer than 20 trials to meet the mastery criterion (12 consecutive correct responses). Figures 4 and 5 (bottom panels) show the pretest and posttest results separately for two categories of relations. As Figure 4 (top) shows, in some cases, the statistical information of the A stimulus supported the hypothesis decision that a student who had completed Experiment 1 (Lesson 2) would reach by attending only to the D stimulus. For example, in Lesson 2 students were taught to *reject the null hypothesis* if an effect corresponded to the scientific-hypothesis prediction. Knowing about a *low p-value* would not change this decision. Such instances will be referred to as *unreversed* relations. As Figure 5 (top) shows, in other cases, the statistical information contradicted the decision that was reinforced in Experiment 1 (Lesson 2) training.

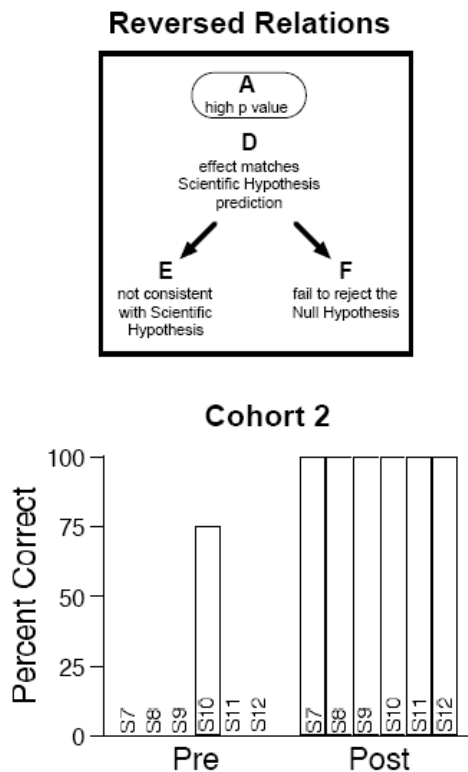


Figure 5

In the example just provided, *high p-value* would dictate *fail to reject the null hypothesis* as the correct answer. Such cases will be referred to as *reversed* relations.

The Experiment 2 tests contained 36 unreversed and 12 reversed relations. If students responded only based on what they learned in Experiment 1 (Lesson 2) -- as would be expected without the benefit of the contextual training -- they would score 100% correct on unreversed relations (Figure 4, bottom) and 0% correct on reversed relations (Figure 5, bottom). On the Experiment 2 pretest, four students (Students 8, 9, 11, and 12) showed precisely this pattern. Student 7 scored 0% correct on reversed relations, as expected, but the addition of statistical information to sample stimuli also disrupted unreversed relations for this student. Curiously, Student 10 showed disruption of unreversed relations and, inconsistent with her performance in the previous lesson, scored 75% correct on reversed relations. We know of nothing unusual about this student that would

account for these results, and were unable to interview her to ask about her strategy.

On the Experiment 2 posttest, all students scored at or near 100% correct on both unreversed and reversed relations. Thus, for four students (Students 8, 9, 11, and 12), Experiment 2 training left intact the abilities that were taught in Experiment 1 (Lesson 2) that applied to unreversed relations (Figure 4, bottom), while establishing reversed relations (Figure 5, bottom). Student 7 also showed clear acquisition of reversed relations and, although this student's unreversed relations initially were disrupted by the addition, in the pretest, of statistical information to the sample stimulus, Experiment 2 training successfully "repaired" these relations. For the remaining participant (Student 10), no claims can be made regarding the effectiveness of the lesson because of unexpectedly high pretest accuracy on reversed relations. In the five other cases, however, Lesson 3 produced clear evidence of acquisition of "conditional reasoning" in hypothesis decision making.

GENERAL DISCUSSION

Among translational applications of stimulus equivalence technology, the present effort is rare in (a) seeking to enhance instruction for advanced learners (see also Ninness, Rumph, McCuller, Harrison, Ford, & Ninness, 2005), and (b) harnessing principles of contextual control. Yet this pilot investigation leaves many important questions unanswered, two of which we identify below.

Unexplored Potential for Emergent Relations

This study documented the reliable emergence of several untaught relations (Figures 2 and 3), but the intervention may have the capacity to generate others that were not evaluated in the Experiment 2 test. This is true because a hallmark of equivalence classes is transfer (or transformation) of function (Dymond & Rehfeldt, 2000). Thus, the contextual-cueing function served by the A stimuli in Experiment 2 (i.e., guiding hypothesis decisions concerning the D stimuli) should have transferred to the B and C stimuli with which these stimuli had become associated in Experiment 1. Figure 6 illustrates the possibilities. Students who were taught to jointly

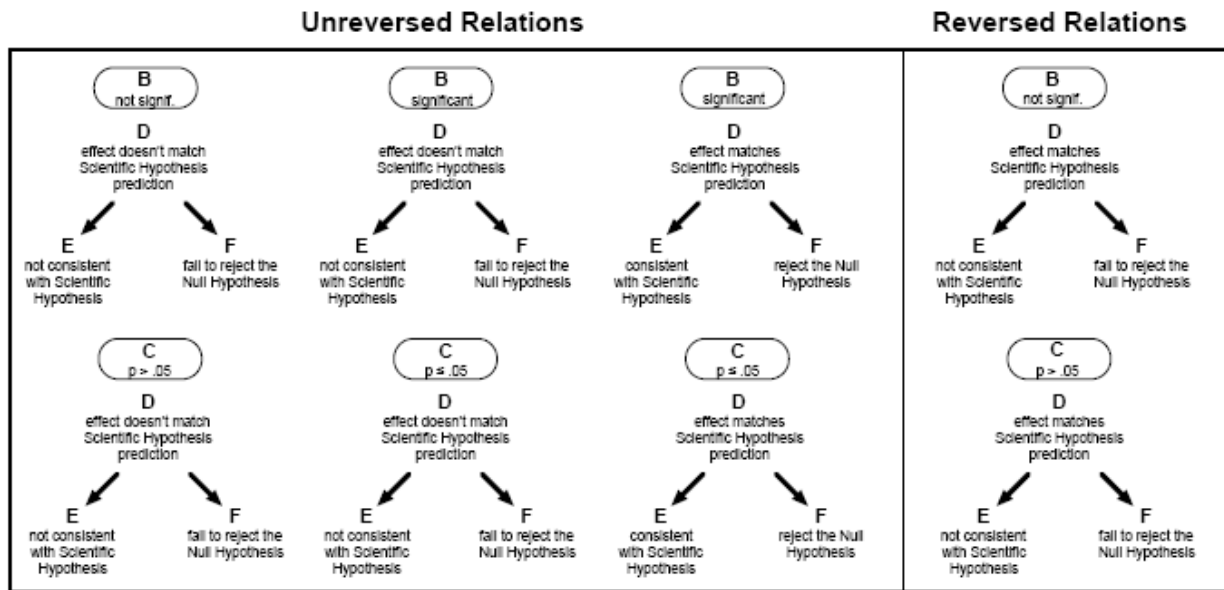


Figure 6

consider D stimuli and A stimuli (*low p-value* and *high p-value*) in rendering hypothesis decisions also should render accurate decisions when shown D stimuli in conjunction with B stimuli (*statistically significant* and *not statistically significant*) or C stimuli ($p \leq .05$ and $p > .05$). Similar transfer of contextual cueing functions has been demonstrated previously in the laboratory (e.g., Gatch & Osborne, 1989; Kohlenberg, et al., 1991) but to date has not received much attention in academic interventions. Unfortunately, we did not test for the emergent relations of Figure 6 due to time constraints, but follow-up work is underway that includes the requisite tests.

A Conceptual Issue: "Purity" of Contextual Control

Procedures designed to establish contextual control over equivalence class membership must jointly present a putative contextual cue with a sample stimulus whose function it purportedly mediates. Although laboratory studies reveal these procedures to be successful in creating conditional class membership (e.g., Bush, et al., 1989; Gatch & Osborne, 1989), as Bush et al. (1989) have pointed out, the resulting effect may be called *contextual control* only under special circumstances. A "pure" contextual cue will regulate class membership for other stimuli without participating in the classes, yet this type

often can be difficult to verify. For example, a sample stimuli and the putative contextual cue might function as a compound sample that is included in some classes but not others; the behavioral outcomes would be indistinguishable from contextual control.

At least three strategies may be imagined for determining whether conditional class membership reflects "pure" contextual control. First, Bush et al. (1989) suggested that contextual cues may have generalized properties; that is, they might regulate class membership for classes other than the ones involved in contextual training. Because the academic material supported only two equivalence classes, this strategy would not be appropriate for the present study. Second, because the components of compound stimuli often develop independent stimulus control over responses to other stimuli in the same equivalence class (e.g., Lane & Critchfield, 1998), it would be useful to present the putative contextual cue (A) alone, in the absence of the sample stimulus (D) whose function it purportedly regulates, to determine whether it is treated as a member of the class in which that sample participates. If it does, then contextual control is suspect. Third, if a stimulus functions as a "pure" contextual stimulus then any other functions that it serves should be not

be affected by this role. In the present investigation, the A stimulus (*low p-value*) was both a putative contextual cue (Experiment 2) and a member of a separate equivalence class (Experiment 1, Lesson 1). It would have been valuable, following the completion of Experiment 2, to test for maintenance of the A-B-C equivalence class from Experiment 1. Disruption of that class would imply, at best, that any contextual function served by *low p-value* in Experiment 2 was not "pure." The critical point is that the two latter means of evaluating "purity" of contextual control could readily be incorporated into an extension of the present translational study -- which is noteworthy because heretofore contextual control has been examined only in a few laboratory investigations, and, of the three types of test mentioned above, only that suggested by Bush et al. (1989) has been attempted (yielding ambiguous results). The present discussion therefore illustrates how translational research, while typically thought of as a mechanism for developing useful applications, also can potentially serve as a vehicle for advancing basic-research agendas (Mace, 1994). In a forthcoming report we will describe results from the third kind of "purity" test described above.

REFERENCES

- Bush, K. M., Sidman, M., & de Rose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior*, 51, 29-45.
- Dixon, M. R., & MacLin, O. H. (2003). *Visual basic for behavioral psychologists*. Reno, NV: Context Press.
- Dymond, S. & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst*, 23, 239-254.
- Gatch, M. B., & Osborne, J. G. (1989). Transfer of contextual stimulus function via equivalence class development. *Journal of the Experimental Analysis of Behavior*, 51, 369-378.
- Huck, S. W. (2000). *Reading statistics and research* (3rd ed.). New York: Longman.
- Knowles, L. (1974). Helping students learn basic inferential statistics. *College Student Journal*, 8, 7-11.
- Kohlenberg, B.S., Hayes, S.C., & Hayes, L.J. (1991). The transfer of contextual control over equivalence classes through equivalence classes: A possible model of social stereotyping. *Journal of the Experimental Analysis of Behavior*, 56, 505-518.
- Lane, S. D., & Critchfield, T. S. (1998). Classification of vowels and consonants by individuals with moderate mental retardation: Development of arbitrary relations via match-to-sample training with compounds. *Journal of Applied Behavior Analysis*, 31, 21-41.
- Mace, F. C. (1994). Basic research needed for stimulating the development of behavioral technologies. *Journal of the Experimental Analysis of Behavior*, 61, 529-550.
- Ninness, C., Rumph, R., McCuller, G., Harrison, C., Ford, A. M., & Ninness, S. K. (2005). A functional analytic approach to computer-interactive mathematics. *Journal of Applied Behavior Analysis*, 38(1), 1-22.
- Sidman, M., & Cresson, O. (1973). Reading and crossmodal transfer of stimulus equivalence in severe retardation. *American Journal of Mental Deficiency*, 77, 515-523.
- Stromer, R. Mackay, H., & Stoddard, L. (1992). Classroom applications of stimulus equivalence technology. *Journal of Behavioral Education*, 2, 225-256.