

*A DISCRIMINATION ANALYSIS OF TRAINING-STRUCTURE
EFFECTS ON STIMULUS EQUIVALENCE OUTCOMES*

RICHARD R. SAUNDERS AND GINA GREEN

SCHIEFELBUSCH INSTITUTE FOR LIFE SPAN STUDIES,
UNIVERSITY OF KANSAS AND
NEW ENGLAND CENTER FOR CHILDREN,
E. K. SHRIVER CENTER FOR MENTAL RETARDATION, AND
NORTHEASTERN UNIVERSITY

Experiments designed to establish stimulus equivalence classes frequently produce differential outcomes that may be attributable to training structure, defined as the order and arrangement of baseline conditional discrimination training trials. Several possible explanations for these differences have been suggested. Here we develop a hypothesis based on an analysis of the simple simultaneous and successive discriminations embedded in conditional discrimination training and testing within each of the training structures that are typically used in stimulus equivalence experiments. Our analysis shows that only the comparison-as-node (many-to-one) structure presents all the simple discriminations in training that are subsequently required for consistently positive outcomes on all tests for the properties of equivalence. The sample-as-node (one-to-many) training structure does not present all the simple discriminations required for positive outcomes on either the symmetry or combined transitivity and symmetry (equivalence) tests. The linear-series training structure presents all the simple discriminations required for consistently positive outcomes on tests for symmetry, but not for symmetry and transitivity combined (equivalence) or transitivity alone. Further, the difference in the number of simple discriminations presented in comparison-as-node training versus the other training structures is larger when the intended class size is greater than three or the number of classes is larger than two. We discuss the relevance of this analysis to interpretations of stimulus equivalence research, as well as some methodological and theoretical implications.

Key words: stimulus equivalence, stimulus classes, simple discrimination, conditional discrimination, discrimination learning, stimulus relations

Experimental analyses of stimulus equivalence based on the Sidman model (1971, 1986, 1994) expose subjects to match-to-sample (MTS) training designed to establish conditional discriminations among stimuli that are not physically similar to one another. Typically, two or more conditional discriminations with some stimuli in common are trained through differential reinforcement. For example, subjects may learn to respond to a comparison stimulus designated B1 if and only if Sample Stimulus A1 is present,

Comparison Stimulus B2 if and only if Sample Stimulus A2 is present, and Comparison Stimulus B3 if and only if Sample Stimulus A3 is present. In addition, subjects may learn to respond to Comparison Stimuli C1, C2, and C3 only in the presence of Sample Stimuli B1, B2, and B3, respectively. If successful, such training establishes conditional relations between each sample (conditional stimulus) and its corresponding correct comparison (discriminative stimulus, or S+), in this case, A1B1, A2B2, A3B3, B1C1, B2C2, and B3C3.

Performances indicating the development of conditional relations are prerequisites for testing the possibility that the trained (or baseline) relations have the properties of equivalence, as defined in mathematics: reflexivity, symmetry, and transitivity. The tests consist of MTS trials on which subjects have the opportunity to demonstrate all possible untrained conditional relations among the stimuli involved in training. Symmetry would be evaluated by reversing the sample and comparison stimuli relative to training (i.e., B1A1, B2A2, B3A3, C1B1, C2B2, C3B3 in our example). Transitivity would be evaluated by

Development of this paper was supported in part by National Institute of Child Health and Human Development Grants HD25995-06 to the E. K. Shriver Center and HD18955-13 and HD02528 to the Schiefelbusch Institute for Life Span Studies, University of Kansas, and by the New England Center for Children. We acknowledge Joe Spradlin and Julie McEntee for their helpful comments on preliminary drafts. Portions of this paper were presented at the third European meeting of the Experimental Analysis of Behaviour Group, July, 1997, in Dublin, Ireland and at the meeting of the Association for Behavior Analysis, May, 1997, Chicago.

Requests for reprints should be addressed to Richard R. Saunders at the Parsons Research Center, 2601 Gabriel, Parsons, Kansas 67357 (E-mail: rrsaun@parsons.lsi.ukans.edu).

presenting sample and comparison stimuli that were related indirectly in training through a common trained relation with other stimuli (i.e., A1C1, A2C2, and A3C3). Tests for the untrained relations C1A1, C2A2, and C3A3 would constitute simultaneous tests for the properties of symmetry and transitivity (also called combined tests, or simply equivalence tests). Reflexivity would be evaluated via conditional identity matching tests with the stimuli involved in training (e.g., A1A1, A2A2, A3A3, B1B1, etc.). If outcomes of all tests are positive, the inference is made that the trained relations are equivalence relations and that the stimuli so related constitute equivalence classes (e.g., Green & Saunders, 1998; R. R. Saunders & Green, 1992; Sidman, 1986, 1994; Sidman et al., 1982; Sidman & Tailby, 1982).

TRAINING STRUCTURE

The foregoing example represents the minimal conditions required to determine whether MTS training produces stimulus equivalence: training on two arbitrary conditional discriminations with one set of stimuli in common (AB and BC in the example), followed by testing for all possible conditional discriminations that were not trained directly. The minimum number of stimuli in an equivalence class is three. In our example, the AB conditional discrimination was trained first, followed by the BC conditional discrimination; the B stimuli were common to both discriminations. Of course, equivalence classes can and often do include more than three stimuli each, and the baseline conditional discriminations can and often are trained in different sequences and with different common stimuli than in our example. The term *training structure* has been used to refer to the sequence of conditional discriminations and the arrangements of common or "linking" stimuli presented to subjects in baseline training. Various terms have been coined to describe specific training structures. For example, Sidman, Kirk, and Willson-Morris (1985) described the situation in which two stimuli are mutually related to a third (e.g., AB, BC) as a "three-stage" training arrangement; training another linked conditional discrimination such as CD created a "four-stage" arrangement (underscored letters designate

common, or linking, sets of stimuli). Fields and Verhave (1987) referred to a stimulus that is related to only one other stimulus in training as a "single," and called a stimulus that is related to more than one other stimulus a "node." To Fields and Verhave, the four-stage arrangement of Sidman et al. (1985) would be a two-node arrangement, with the B and C stimuli serving as nodes. These authors also suggested that the structure of equivalence classes can be described in terms of four parameters: the number of stimuli in each class, the number of nodes, the pattern of singles relative to nodes, and the pattern formed by assignment of stimuli to the roles of samples and comparisons during training (i.e., directionality of training).

The original Sidman analysis of stimulus equivalence did not suggest that equivalence test outcomes should vary as a function of training structure, order, or direction (Sidman & Tailby, 1982). On the contrary, it implied that if training established the intended conditional relations and concurrently prevented the development of extraneous stimulus control, then responding on all tests for untrained relations should be consistent with equivalence, regardless of the order and arrangement of the trained conditional discriminations (Carrigan & Sidman, 1992; Green & Saunders, 1998; Sidman, 1994). Some investigators, however, have reported differential outcomes on equivalence tests that appear to be due to training structure. For example, one recent study with preschool children found that five-member equivalence classes were more likely following one training sequence than another (R. R. Saunders, Drake, & Spradlin, 1999). All of the children were exposed to two-choice MTS training designed to establish four conditional discriminations among 10 arbitrary visual stimuli. For 6 children, two stimuli served as the samples in all four conditional discriminations; the trained relations were designated AB, AC, AD, and AE. This kind of training structure has been referred to as a "sample-as-node" or "one-to-many" structure. Five other children received training with one pair of stimuli serving as comparisons with each of four different pairs of sample stimuli. The trained conditional relations were designated BA, CA, DA, and EA. This training structure has been dubbed a "comparison-as-node" or

“many-to-one” structure (cf. K. J. Saunders, Saunders, Williams, & Spradlin, 1993; Urcuioli & Zentall, 1993; Urcuioli, Zentall, Jackson-Smith, & Steirn, 1989). Positive outcomes on equivalence tests were found for all 5 children who had comparison-as-node training, but positive outcomes were found for only 2 of 6 children who had sample-as-node training.

The results obtained by R. R. Saunders et al. (1999) replicated those of previous experiments with adolescents and adults with mental retardation (Drake & Saunders, 1987, cited in K. J. Saunders et al., 1993; R. R. Saunders, Wachter, & Spradlin, 1988; Spradlin & Saunders, 1986). Across these studies, five-member equivalence classes were established in only 1 of 7 subjects trained with the sample-as-node procedure, but they were established in 6 of 6 subjects trained with the comparison-as-node procedure. Apparent structure-related differences have also been reported with normally capable adult subjects (Barnes, 1992, as cited in Barnes, 1994; Fields, Hobbie, Adams, & Reeve, in press). In contrast, a recent study with normally capable adults suggested that sample-as-node training was more likely to produce three-member equivalence classes than was comparison-as-node training (Arntzen & Holth, 1997). Finally, some investigators have reported negative outcomes on tests for some properties of equivalence following baseline training in which several conditional discriminations were trained in sequence, with multiple nodal or linking stimuli (e.g., AB, BC, CD, DE; see Arntzen & Holth, 1997; Fields, Landon-Jimenez, Buffington, & Adams, 1995; Holth & Arntzen, 1998). This has been described as a linear-series training structure (e.g., Green & Saunders, 1998).

SOME HYPOTHESES ABOUT TRAINING-STRUCTURE EFFECTS

Why might different training structures yield different outcomes on equivalence tests? Several possibilities have been suggested. With respect to linear-series training structures, Fields and colleagues offered an account based on nodal or “associative” distance. They suggested that the larger the number of nodes potentially linking stimuli indirectly in training, the less robust the per-

formances on tests for the untrained relations among those stimuli were likely to be (e.g., Fields, Verhave, & Fath, 1984). Although the results of some experiments seem to support this hypothesis (e.g., Fields, Adams, Verhave, & Newman, 1990; Kennedy, 1991; Kennedy, Itkonen, & Lindquist, 1994), procedural variations and stimulus characteristics in those experiments make them difficult to interpret. (These will be discussed later.) This account has also been criticized because it invokes labels (nodal distance, associative distance) for a structural property (the number of nodes) that may imply that other hypothetical structural properties are at work (Sidman, 1994, p. 539). Further, the account falls short of explaining different outcomes on tests for equivalence in basic behavioral terms.

With respect to sample-as-node versus comparison-as-node differences, Sidman (1994, pp. 527–528) raised the possibility that the former might establish differential contextual control of trained conditional relations by negative stimuli (incorrect comparisons), which could lead to negative outcomes on tests for the properties of equivalence. An alternative account was postulated by Spradlin and his colleagues (K. J. Saunders et al., 1993; Spradlin & Saunders, 1986; also see Sidman, 1994, pp. 526–527). They speculated that sample-as-node training did not consistently produce equivalence classes because the training contingencies did not require all the simple discriminations subsequently called for on tests for equivalence. Spradlin and colleagues noted that when AB and AC are trained, for example, subjects need only discriminate the A samples from one another, the B comparisons from one another, and the C comparisons from one another; the training contingencies do not require discrimination of each B stimulus from each C stimulus. The B versus C discriminations are called for, however, on the BC and CB trials that constitute tests for the properties of equivalence. In contrast, comparison-as-node training (e.g., BA and CA) requires successive discrimination of all B and C stimuli across trials and the simultaneous discrimination of all A comparisons within trials to fulfill the training contingency requirements. That is, this training structure potentially establishes all of the simple discriminations required for consistently positive outcomes on tests for the properties of equiv-

alence (K. J. Saunders et al., 1993; R. R. Saunders et al., 1999; Spradlin & Saunders, 1986; also see Barnes, 1994; Sidman, 1994).

THE DISCRIMINATION ACCOUNT: AN ELABORATION AND EXPANSION

We propose that, with some additional development and elaboration, the discrimination analysis suggested by Spradlin and colleagues provides a parsimonious account of the differential effects of training structures on equivalence test outcomes, one that is consistent with basic principles of stimulus control and does not invoke constructs like associative distance or mediation (cf. McIlvane & Dube, 1992; Sidman, 1986, 1994). Here we undertake such an elaboration by examining (a) how simple simultaneous and successive discriminations are necessarily embedded in conditional discriminations; (b) which and how many of those component simple discriminations are presented to subjects in each of the training structures commonly used in stimulus equivalence experiments; and (c) the component simple discriminations required for consistently positive outcomes on tests for equivalence following training with each structure. Finally, we reanalyze the results of several published studies of stimulus equivalence from this perspective.

Assumptions and Definitions

For our analysis, we adopt a critical assumption that has been stated or implied by several other authors (e.g., McIlvane & Dube, 1996; K. J. Saunders et al., 1993; Spradlin & Saunders, 1986; Sidman, 1994): *For performances to meet criteria for acquisition of the trained baseline relations as well as criteria for positive outcomes on all tests for stimulus equivalence, each stimulus must be discriminated from every other stimulus in the experiment.* On its face, this assumption might appear counterintuitive. It might seem that in order to “pass” all tests for stimulus equivalence, subjects need only discriminate all stimuli in each class from all stimuli in the other classes (between-class discriminations, such as A1 vs. A2 vs. A3, B1 vs. B2 vs. B3), and need not discriminate stimuli within classes from one another (A1 vs. B1, B2 vs. C2, C3 vs. D3, etc.). The reasoning might go like this: If training contingencies

establish that two sample stimuli (e.g., B1 and C1) both control a response to the same comparison (e.g., A1), might not the subject then treat B1 and C1 as if they are the same—that is, fail to discriminate them? And would this not suffice to produce apparently positive outcomes on all tests for equivalence (e.g., trials testing B1C1 and C1B1), as long as the subject discriminates B1 and C1 from the stimuli in the other prospective equivalence classes? We maintain that the answer to these questions is no, for the following reasons: At the beginning of a well-designed stimulus equivalence experiment, the subject is exposed to a group of unsorted stimuli that do not bear any consistent physical resemblance to one another. The initial training contingencies are designed to establish a relation between each comparison stimulus and a particular sample stimulus. Therefore, the contingencies require discrimination of every sample from every other sample presented successively across trials, all samples from all comparisons, and all comparison stimuli presented simultaneously within trials (Sidman, 1986). Training procedures that enhance the probability that every stimulus will be discriminated from every other stimulus—such as contingencies that specify a different response to each conditional (sample) stimulus and to each discriminative (comparison) stimulus—should foster the development of conditional discriminations and, therefore, the development of equivalence classes (Sidman, 1994, pp. 413–414).

Like training trials, tests for stimulus equivalence also present conditional discriminations that are composed of simple successive and simultaneous discriminations among experimental stimuli. Therefore, for consistently positive outcomes on all tests for the properties of equivalence, every stimulus must be discriminated from every other stimulus. More important, however, test trials may present some simple successive and simultaneous discriminations that were not presented at all in training, or that are presented differently on tests than in training (e.g., successively vs. simultaneously). If the relevant simple discriminations that compose the tested conditional discriminations are not made, negative outcomes on some or all tests for the properties of equivalence will result (see McIlvane & Dube, 1996; K. J. Saunders et al., 1993; R.

R. Saunders et al., 1999; Sidman, 1986, 1994; Spradlin & Saunders, 1986). We argue here that different training structures make such outcomes more or less probable because they are more or less likely to establish the necessary simple discriminations.

To bolster our contention that consistently positive outcomes on tests for stimulus equivalence require discrimination of every stimulus from every other stimulus in the experiment, it is important to note that although such outcomes imply that the stimuli within classes are *substitutable* for one another, it does not necessarily follow that they are *indiscriminable* from one another. On the contrary, from a practical standpoint it is vital for organisms to discriminate that stimuli that can substitute for one another in certain contexts nonetheless have distinctive features. For example, we may observe that a child matches an apple, a picture of an apple, and the printed word APPLE to one another in every possible sample-comparison arrangement. That observation might lead us to conclude that those stimuli are substitutable for one another, and constitute an equivalence class. But that observation should not lead to the conclusion that those stimuli are not discriminated. Indeed, for the class to be functional for the child, it is necessary for discriminations among the members to be *maintained*—so that the child does not try to eat the picture or the printed word, for example—at the same time as the stimuli are treated as equivalent (substitutable) in certain contexts. (Whether discriminations among stimuli are in fact maintained during or after demonstrations that they are equivalent could be evaluated empirically via conditional identity matching tests; cf. Dube, McIlvane, & Green, 1992.) Of course, it is also necessary for all members of the “apple” class to be discriminated from all members of other stimulus classes. In other words, performances that consistently demonstrate stimulus equivalence entail discriminations among stimuli within as well as between classes.

For purposes of our analysis, we refer to the grand total of all possible simple simultaneous and successive discriminations among the stimuli in a stimulus equivalence experiment as *required* for consistently positive outcomes on all training and test trials.

We describe the types and proportions of those discriminations that are *presented* and *not presented* in each of the training structures typically used in stimulus equivalence experiments, to reflect the understanding that whether the training contingencies actually establish those discriminations is always an empirical question.

Our analysis also assumes (a) the use of simultaneous MTS procedures; (b) that baseline conditional discriminations are mixed at some point before testing for equivalence begins (e.g., if AB and AC relations are trained, all of those trial types are presented together within one or more training session before testing); (c) that all trial types testing for a particular property of equivalence (e.g., BC and CB) are presented together within the same session; and (d) that training and testing are conducted with balanced MTS conditional discrimination procedures (Green & Saunders, 1998). For example, Comparison C1 is always presented with Comparisons C2 and C3; C1 is never presented as a comparison with any stimuli from other sets (e.g., B2, D3). A final assumption is that no differential consequences are arranged for responses on test trials.

Simple Discriminations Embedded in Conditional Discriminations

We turn now to a review of simple and conditional discrimination training contingencies, to reiterate precisely why acquisition of conditional discriminations necessitates acquisition of the component simple discriminations.

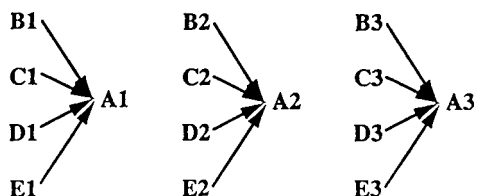
Simple discriminations. In the presence of a particular antecedent stimulus, if a particular response is followed by a particular consequence, the response may come to occur more often in the presence of the stimulus (S+) than in its absence or in the presence of another stimulus (S−). When repeated application of these contingencies leads to responding in the presence of the S+ but not in the presence of the S−, one can infer that a simple discrimination has been established. When the S+ and the S− are presented concurrently on each of a series of trials, the procedure is termed a *simultaneous simple discrimination*. When the stimuli are presented one at a time, on unsystematically alternating tri-

als, the procedure is termed a *successive simple discrimination*.

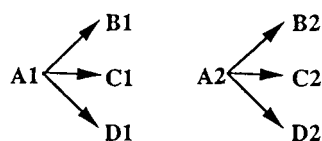
Conditional discriminations. In conditional discrimination training, simple discriminations are brought under the control of additional antecedents, or conditional stimuli (Sidman, 1986). A common and reliable preparation for establishing conditional discriminations is MTS training. The MTS procedures used in most stimulus equivalence experiments involve a minimum of two different conditional stimuli (samples) and two different discriminative stimuli (comparisons) per conditional discrimination. The same comparison stimuli are presented on every trial while the sample stimulus varies unsystematically from trial to trial. In the least complicated procedure, contingencies are arranged so that each comparison stimulus is discriminative for reinforcement (S+) in the presence of one and only one conditional (sample) stimulus, and is not discriminative for reinforcement (S-) in the presence of a second sample stimulus. That is, the functions of the discriminative stimuli change from trial to trial, depending on which sample stimulus is present. Although three or more samples and comparisons are generally desirable (Carrigan & Sidman, 1992; Green & Saunders, 1998; Johnson & Sidman, 1993; Kelly, Green, & Sidman, 1998; Sidman, 1987), we use just two samples and comparisons here and elsewhere for ease of illustration.

In a typical stimulus equivalence experiment, multiple training trials are presented in a session or block. A different sample is presented on each trial, and the same comparisons are presented on every trial. Each sample is usually presented equally often within a session or block of trials, in unsystematic order. The positions of the comparisons, especially the S+, vary unsystematically from trial to trial. Sample stimuli are presented one at a time across trials, with some time elapsing between presentations; thus, discriminating among them involves successive simple discriminations. Comparison stimuli are presented together on every trial, as are the samples and comparisons that constitute each trial type; thus, discriminating among them involves simultaneous simple discriminations. In short, performing conditional discriminations necessarily involves a number of simple discriminations, some simultaneous,

Three-Choice Comparison-as-Node



Two-Choice Sample-as-Node



Four-Choice Linear Series

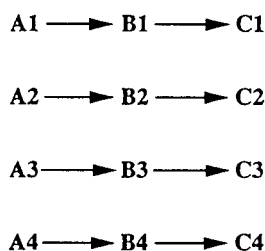


Fig. 1. Representations of three basic training structures used in stimulus equivalence experiments: comparison as node (upper panel) with three potential classes of five stimuli each; sample as node (middle panel) with two potential classes of four stimuli each; and linear series (lower panel) with four potential classes of three stimuli each. Arrows point from stimuli used as samples to comparison stimuli.

some successive (cf. Green & Saunders, 1998; K. J. Saunders & Spradlin, 1989, 1993; Sidman, 1986).

Common Stimulus Equivalence Training Structures

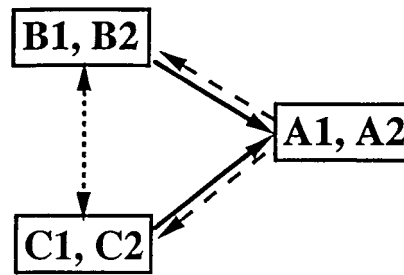
Many variations of the training structures described by Fields and Verhave (1987) appear in the stimulus equivalence literature, but there are three basic prototypes. The top panel of Figure 1 shows a comparison-as-node (many-to-one) structure for potentially developing three equivalence classes with five stimuli per class. This training structure usually involves the use of three-choice MTS proce-

dures. The middle panel of Figure 1 shows the sample-as-node (one-to-many) structure for potentially developing two equivalence classes with four stimuli per class. The training represented is usually arranged using two-choice MTS procedures. The linear-series structure is shown in the lower panel of Figure 1. The schematic depicts four potential classes of three stimuli each, which usually involves four-choice MTS procedures. Most stimulus equivalence experiments in the literature have used comparison-as-node, sample-as-node, or linear-series training structures, or variations thereof, to produce two or more equivalence classes (Green & Saunders, 1998; for some examples of variations, see Fields, Adams, & Verhave, 1993; Kennedy, 1991; Pilgrim & Galizio, 1995; Spradlin, Cotter, & Baxley, 1973; Wetherby, Karlan, & Spradlin, 1983).

Simple Discriminations Presented in Various Training Structures

Comparison as node. Figure 2 represents comparison-as-node training (BA and CA) leading potentially to the development of two equivalence classes of three stimuli each. Trial types for this training, and for tests of symmetry and equivalence, are shown beneath the schematic. In this structure, 15 simple discriminations are presented in training, as follows: Comparison Stimuli A1 and A2 are presented simultaneously on training trials (one simple discrimination). Each comparison stimulus is presented simultaneously with each sample stimulus; that is, A1 is presented with B1, B2, C1, and C2 (four discriminations), and A2 is presented with B1, B2, C1, and C2 (four discriminations). The two B stimuli and the two C stimuli are presented successively across training trials as samples when BA and CA trials are mixed (six discriminations). For the training contingency requirements to be fulfilled consistently, all 15 of these simple discriminations must be made.

Tests for stimulus equivalence—BC and CB trial types—also present certain simple discriminations as components of the tested conditional discriminations: the successive discrimination of all B and C sample stimuli, two simultaneous discriminations between the comparison stimuli (B1 vs. B2 and C1 vs. C2), and four simultaneous discriminations



Trial Types

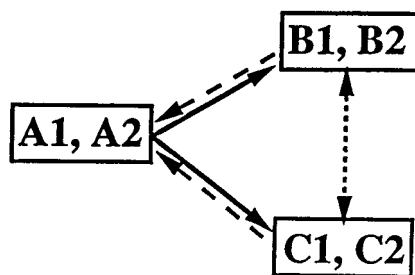
	Samples	Comparisons
Training	B1	*A1 A2
	B2	A1 *A2
	C1	*A1 A2
	C2	A1 *A2
Symmetry	A1	*B1 B2
	A2	B1 *B2
	A1	*C1 C2
	A2	C1 *C2
Equivalence	B1	*C1 C2
	B2	C1 *C2
	C1	*B1 B2
	C2	B1 *B2

Fig. 2. Comparison-as-node training to produce two three-member equivalence classes. Solid arrows indicate trained relations, dashed-line arrows indicate tests for the property of symmetry, and the double-headed dotted-line arrows indicate combined tests for the properties of symmetry and transitivity (or equivalence tests). Trial types for training trials and tests for symmetry and equivalence are shown below the schematic. Asterisks indicate experimenter-designated correct comparisons (for positive test outcomes).

between samples and comparisons (i.e., each B stimulus from each C stimulus). In this case, all of those simple discriminations were presented during baseline training. Although the B and C stimuli were not presented as simultaneous discriminations, they were presented as successive (sample) discriminations when the BA and CA training trials were mixed prior to testing. Because simultaneous discriminations are generally easier than successive discriminations (Brady & Saunders, 1991; Carter & Eckerman, 1975; Urcuioli et

al., 1989), we would expect most subjects to have little difficulty with the BC simultaneous discriminations called for on the equivalence tests in this example. Similarly, all of the simple discriminations involved in tests for symmetry (AB and AC) were presented during training. Only one (A1 vs. A2) might pose a problem for some subjects, because the discrimination between those stimuli was presented simultaneously in training, but it is presented successively on the symmetry tests. Most of the other simple discriminations called for on the symmetry tests are exactly the same as in training; two (B1 vs. B2 and C1 vs. C2) that were presented successively in training appear as simultaneous discriminations on symmetry tests. The discriminations between the A stimuli and the B and C stimuli remain the same as in training, except that sample and comparison roles are reversed. In short, training with this comparison-as-node structure presents all of the simple discriminations involved in every symmetry and equivalence test trial type. (Tests for transitivity alone are not possible following training with this structure.)

Sample as node. Quite a different picture emerges from an analysis of the simple discriminations presented during sample-as-node training. Figure 3 represents sample-as-node training (AB and AC) leading potentially to the development of two equivalence classes of three stimuli each. In training, Comparison Stimuli B1 and B2 are presented simultaneously, as are Comparisons C1 and C2. Samples A1 and A2 are presented successively, and are presented simultaneously with the B and C comparisons. Although the B and C stimuli are presented in pairs successively across trials when AB and AC training trials are mixed, the B and C stimuli are never pitted against one another within a trial (see Barnes, 1994), nor are they presented successively as samples. Thus, successful training with this sample-as-node structure potentially results in every stimulus being discriminated from every other stimulus except the B stimuli from the C stimuli. Discriminations among the B and C stimuli are first called for either on tests for equivalence (BC, CB) or on tests for symmetry (BA, CA), whichever is presented first. As the lower portion of Figure 3 shows, all four B and C stimuli must be discriminated from one another



Trial Types

	Samples	Comparisons
Training	A1	*B1 B2
	A2	B1 *B2
	A1	*C1 C2
	A2	C1 *C2
Symmetry	B1	*A1 A2
	B2	A1 *A2
	C1	*A1 A2
	C2	A1 *A2
Equivalence	B1	*C1 C2
	B2	C1 *C2
	C1	*B1 B2
	C2	B1 *B2

Fig. 3. Sample-as-node training to produce two three-member equivalence classes. Solid arrows indicate trained relations, dashed-line arrows indicate tests for the property of symmetry, and the double-pointed dotted-line arrows indicate combined tests for the properties of symmetry and transitivity (or equivalence tests). Trial types for training trials and tests for symmetry and equivalence are shown below the schematic. Asterisks indicate experimenter-designated correct comparisons (for positive test outcomes).

as samples across trials on the symmetry tests, and as samples across trials as well as samples and comparisons within trials on the tests for equivalence. On symmetry tests, the simultaneous discrimination of A1 from A2 as comparisons is not likely to be problematic, because those stimuli were presented in successive discrimination format in training. In summary, negative results on equivalence and symmetry tests are expected to be more likely following sample-as-node training, because sam-

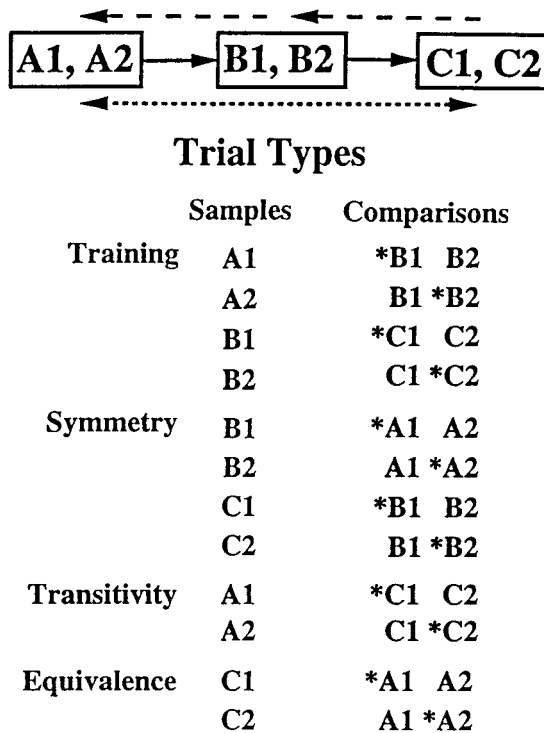


Fig. 4. Linear-series training to produce two three-member equivalence classes. Solid arrows indicate trained relations, dashed-line arrows indicate tests for the property of symmetry, and the double-pointed dotted-line arrow indicates the test for the properties of transitivity (AC) and symmetry and transitivity combined, or equivalence (CA). Trial types for training trials and tests for symmetry, transitivity, and equivalence are shown below the schematic. Asterisks indicate experimenter-designated correct comparisons (for positive test outcomes).

ple-as-node training does not present some of the simple discriminations that make up the conditional discriminations called for on the tests.

Linear series. Examination of the simple discriminations involved in linear-series training, shown in Figure 4, reveals yet another pattern. In AB and BC training to establish two classes of three stimuli each, the B stimuli are presented simultaneously as comparisons, as are the C stimuli. The A and B stimuli are presented simultaneously as samples and comparisons on the AB training trials, and the B and C stimuli are presented simultaneously on the BC training trials. The A and B stimuli are presented successively as samples when the AB and BC training trials are mixed. On the symmetry tests (BA and CB),

the B and C stimuli are presented successively as samples; they were presented simultaneously in training. As noted previously, this might pose a problem for some subjects. In addition, the CB symmetry test trials call for the successive discrimination of C1 from C2, which was never presented in training, whereas the BA symmetry test trials involve successive (B1 vs. B2) as well as simultaneous (A1 vs. A2, B stimuli vs. A stimuli) discriminations that were presented in training. Thus, different outcomes may occur on these two types of symmetry test trials. Note that with linear-series training designed to establish larger equivalence classes (e.g., AB, BC, CD, DE), only the test for symmetry of the final conditional discrimination (e.g., ED) will suffer from the potential problem just described for CB test trials.

Referring to the lower portion of Figure 4, it is evident that the tests for transitivity and equivalence following linear series training present both simple and successive discriminations among stimuli (A and C) that are never pitted against one another in training. The only context in which those stimuli are presented together in training is when the AB and BC training trials are mixed; there the A and C stimuli are presented successively across trials within a session, but as samples and comparisons, respectively, on different trial types. Thus, as with sample-as-node training structures, linear-series structures are expected to result in a higher probability of failure on equivalence tests than training with comparison-as-node structures.

Simple Discriminations in Larger Training Structures

Table 1 summarizes the preceding analysis. To reiterate, in any training structure designed to produce two equivalence classes of three stimuli each, the grand total of possible simple simultaneous and successive discriminations among all the stimuli is 15. The third column of the table contrasts the proportion of those discriminations *not presented* in each of the training structures. It shows that all of the potential simple discriminations are presented in comparison-as-node training, meaning that the subject encounters no novel discriminations on tests for equivalence and symmetry. Training with the other structures, however, omits some of the simple discrimi-

Table 1

Simple discriminations in each of three training structures designed to produce two three-member equivalence classes.

Structure	Grand total	Number not presented in training/total	Number not presented in training, but presented on equivalence tests	Number not presented in training, but presented on symmetry tests
Comparison as node	15	0/15	0	0
Sample as node	15	4/15	4	4
Linear series	15	4/15	4	0

nations that are presented on tests for equivalence, as indicated by the entries in the fourth column. The entries in the fifth column show that the discriminations not presented in sample-as-node training are presented in tests for symmetry. This is not the case for linear-series training.

Table 2 compares the numbers of simple discriminations presented in the three training structures when training is designed to produce two equivalence classes of four stimuli each. Comparing the entries in Table 3 with those in Table 2 reveals that increasing class size from three to four stimuli increases the total number of simple discriminations from 15 to 28. It also changes the proportion of those discriminations that are presented in training, as well as the proportion of simple discriminations not presented in training but called for on tests for the properties of equivalence. In comparison-as-node training, all the simple discriminations among the stimuli are presented in training. In sample-as-node and linear-series training structures, 12 simple discriminations are not presented in training. Following sample-as-node training,

all of those discriminations are presented on both equivalence and symmetry tests. Following linear-series training, all of the simple discriminations not presented in training are presented on equivalence tests, but none of them are presented on symmetry tests.

Table 3 shows the effects of holding prospective class size at four stimuli per class while increasing the number of prospective classes to three and employing three-choice MTS training procedures (to insure balanced trial types per the assumptions underpinning our analysis). The total number of possible discriminations increases to 66. As in the preceding examples, comparison-as-node training presents all of them. In contrast, increasing the number of prospective equivalence classes from two to three alters the proportion of the total simple discriminations that are presented in sample-as-node and linear-series training structures. As Table 3 indicates, 27 of the possible simple discriminations among the experimental stimuli are not presented in sample-as-node and linear-series training; all of those discriminations appear on the equivalence tests that follow training

Table 2

Simple discriminations in each of three training structures designed to produce two four-member equivalence classes.

Structure	Grand total	Number not presented in training/total	Number not presented in training, but presented on equivalence tests	Number not presented in training, but presented on symmetry tests
Comparison as node	28	0/28	0	0
Sample as node	28	12/28	12	12
Linear series	28	12/28	12	0

Table 3

Simple discriminations in each of three training structures designed to produce three four-member equivalence classes.

Structure	Grand total	Number not presented in training/total	Number not presented in training, but presented on equivalence tests	Number not presented in training, but presented on symmetry tests
Comparison as node	66	0/66	0	0
Sample as node	66	27/66	27	27
Linear series	66	27/66	27	0

with both of those structures. Those 27 discriminations are also presented on symmetry tests following sample-as-node, but not linear-series, training.

Table 4 shows the complete map of the simple discriminations summarized in Table 3 for sample-as-node training. This table distinguishes the simple discriminations that are and are not presented in training, and reflects the roles of the stimuli (i.e., samples, comparisons) involved in each simple discrimination. A similar map for linear-series training would be organized differently, but

the category totals would be the same. For example, the discrimination of A1 from C3 would not be presented in linear-series training, but the discrimination of B1 from C3 would be.

As the series of preceding tables suggests, the differences between the proportion of simple discriminations presented in comparison-as-node training and those presented in the other structures increase as prospective class size and number of classes increase. Table 5 shows that for sample-as-node and linear-series training, the number of simple dis-

Table 4

Simple discriminations in sample-as-node training with 12 stimuli and three conditional discriminations (AB, AC, AD) to produce three four-member equivalence classes.

Discrimination	Stimuli involved	Simple discriminations		
Presented in training				
Successive	Samples	A1 vs. A2	A1 vs. A3	A2 vs. A3
Simultaneous	Samples and comparisons	A1 vs. B1	A1 vs. B2	A1 vs. B3
		A1 vs. C1	A1 vs. C2	A1 vs. C3
		A1 vs. D1	A1 vs. D2	A1 vs. D3
		A2 vs. B1	A2 vs. B2	A2 vs. B3
		A2 vs. C1	A2 vs. C2	A2 vs. C3
		A2 vs. D1	A2 vs. D2	A2 vs. D3
		A3 vs. B1	A3 vs. B2	A3 vs. B3
		A3 vs. C1	A3 vs. C2	A3 vs. C3
		A3 vs. D1	A3 vs. D2	A3 vs. D3
	Comparisons	B1 vs. B2	B1 vs. B3	B2 vs. B3
		C1 vs. C2	C1 vs. C3	C2 vs. C3
		D1 vs. D2	D1 vs. D3	D2 vs. D3
	Not presented in training			
Comparisons	Comparisons	B1 vs. C1	B1 vs. C2	B1 vs. C3
		B1 vs. D1	B1 vs. D2	B1 vs. D3
		B2 vs. C1	B2 vs. C2	B2 vs. C3
		B2 vs. D1	B2 vs. D2	B2 vs. D3
		B3 vs. C1	B3 vs. C2	B3 vs. C3
		B3 vs. D1	B3 vs. D2	B3 vs. D3
		C1 vs. D1	C1 vs. D2	C1 vs. D3
		C2 vs. D1	C2 vs. D2	C2 vs. D3
		C3 vs. D1	C3 vs. D2	C3 vs. D3

Table 5

Simple discriminations by class size and number of classes for linear series and sample-as-node training structures.

Stimuli per class	Number of classes	Grand total discriminations	Number of discriminations not presented in training
3	2	15 ^a	4 ^a
	3	36	9
	4	66	16
	5	105	25
4	2	28 ^b	12 ^b
	3	66 ^c	27 ^c
	4	120	48
	5	190	75
5	2	45	24
	3	105	54
	4	190	96
	5	300	150
6	2	66	40
	3	153	90
	4	276	160
	5	435	250

^a From Table 1.

^b From Table 2.

^c From Table 3.

criminations not presented in training increases as a direct function of class size and class number. Moreover, given a particular number of classes, the proportion of discriminations not presented to the total number of discriminations also increases as class size is increased. The proportion is relatively unaffected by increases in number of classes, however, when class size is held constant. Whether increases in the proportion of discriminations not presented will have different effects than increases in number alone is a question for future research.

In sum, sample-as-node or linear-series training structures for large equivalence classes present only a fraction of the simple discriminations that make up the conditional discriminations encountered on tests for the properties of equivalence. It is possible that some new discriminations will be acquired over the course of repeated testing, which is often conducted when positive outcomes are not seen on initial tests (see R. R. Saunders & Green, 1992). Because these additional discriminations must be acquired in the absence of trial-by-trial differential consequences, they are not likely to be acquired rapidly, if they

are acquired at all. This may account for the gradual emergence of equivalence that has been reported by a number of investigators, a point to which we will return later.

APPLICATION OF THE DISCRIMINATION ANALYSIS TO EQUIVALENCE RESEARCH

The analysis presented here suggests some testable hypotheses about the results of stimulus equivalence experiments conducted with various training structures. The most general one is that, all other things being equal, sample-as-node and linear-series training are less likely to yield positive outcomes on tests for equivalence than is comparison-as-node training. In the following sections we examine relevant published studies for evidence bearing on this and related hypotheses.

Studies Involving a Small Number of Small Classes

Investigators have reported mixed outcomes on tests for equivalence in young children following either sample-as-node training (Barnes, Browne, Smeets, & Roche, 1995; Barnes, McCullagh, & Keenan, 1990; Devany, Hayes, & Nelson, 1986; Pilgrim, Chambers, & Galizio, 1995; Sidman et al., 1985; Sidman, Willson-Morris, & Kirk, 1986) or linear-series training (Lazar, Davis-Lang, & Sanchez, 1984; Michael & Bernstein, 1991) to establish two three-member classes. In an experiment with 3 children aged 6 to 7 years, a sample-as-node structure was used to establish three three-member classes of tactile stimuli. Equivalence classes were eventually established, but only after extensive retesting (O'Leary & Bush, 1996). Another study, using comparison-as-node training to establish three three-member classes of arbitrary visual stimuli with children aged 7 to 12 years, produced positive test outcomes immediately following training (Williams, Saunders, Saunders, & Spradlin, 1995). In contrast, Eikeseth and Smith (1992) reported that two three-member classes were established in none of 4 young children with autism following sample-as-node training. It is important to note that the objective of the initial testing and training in all of these experiments was to establish just two or three three-member equivalence classes, so that the

total number of simple discriminations involved was quite small.

Most experiments using sample-as-node and comparison-as-node training to establish small equivalence classes with adults have produced positive outcomes. Green (1990) reported development of two auditory-visual as well as two all-visual equivalence classes of three members each in adults with mild mental retardation following sample-as-node training. In another study using a sample-as-node structure, three three-member classes were established in 15 of 16 normally capable adults (Innis, Lane, Miller, & Critchfield, 1998). Pilgrim and Galizio (1990) also reported positive outcomes in normally capable adults following sample-as-node training. More recently, the same investigators reported establishing two four-member equivalence classes in normally capable adults using a mixed comparison-as-node/sample-as-node training structure (Pilgrim & Galizio, 1995).

Investigators using linear-series training structures designed to establish small equivalence classes have reported mixed results. Fields, Adams, Newman, and Verhave (1992) reported that immediately after linear-series training, positive outcomes on tests for all properties of equivalence were evident for only 3 of 14 adults; repeated or specially designed testing was necessary to yield positive outcomes in the other subjects. Similar results were obtained in the same laboratory with other adult subjects, and successful linear-series training and testing with one set of stimuli was not always followed by positive outcomes following linear-series training with a second set of stimuli (Adams, Fields, & Verhave, 1993; Buffington, Fields, & Adams, 1997; Fields et al., 1997). The studies by Fields and colleagues employed consonant-vowel-consonant trigrams as stimuli. In contrast, three-choice linear-series training with olfactory stimuli, trigrams, and arbitrary forms (Annett & Leslie, 1995) and a mixture of trigrams and haptic stimuli (Tierney, de Lary, & Bracken, 1995) produced positive results in most adult subjects.

In summary, the observation that sample-as-node and linear-series training established equivalence classes in some subjects suggests that the successive discriminations that were called for on the tests for equivalence were established, even though they were not ex-

PLICITLY presented in baseline training. We noted previously that although sample-as-node training (e.g., AB, AC) does not explicitly require discrimination of comparison stimuli from different trial types (e.g., B stimuli from C stimuli) because those stimuli are never pitted directly against one another, they are presented successively across trials as pairs of comparisons when training trials are mixed. With small potential classes, this exposure may suffice for acquisition of the B versus C discriminations by some subjects. The same possibility applies to linear-series training for a small number of small classes. For example, the A stimuli may be discriminated from the C stimuli when AB and BC trials are mixed during training. If so, positive test outcomes would seem equally likely following training with either structure when the total number of simple discriminations involved is relatively small. Alternatively, positive outcomes might result from interspersing test trials with training trials in test sessions. Previously untrained simple discriminations could develop over the course of testing due to (a) the additional exposure to training trials, (b) the juxtaposition of test trials that include those discriminations with training trials, or (c) both. This might explain the gradual emergence of equivalence-consistent test performances that has been documented in a number of studies (e.g., Lazar et al., 1984; Sidman et al., 1986; Spradlin et al., 1973).

Studies Involving Larger Classes

With sample-as-node or linear-series training designed to produce more than two potential classes or classes with more than three members, the number of new discriminations presented to subjects on tests for equivalence is greater than with smaller classes (refer to Table 5). Therefore, we would expect structure-related differences in test performances to be more likely or more marked when training is designed to produce larger classes. Results of a series of studies by R. R. Saunders and colleagues, discussed previously, are consistent with this prediction: In 23 of 28 young children and individuals with mental retardation, four- and five-member equivalence classes were established following comparison-as-node training; in only 3 of 13 subjects were such classes established following sample-as-node training (Drake & Saunders,

1987, cited in K. J. Saunders et al., 1993; K. J. Saunders et al., 1993; R. R. Saunders et al., 1999; R. R. Saunders, Saunders, Kirby, & Spradlin, 1988; R. R. Saunders, Wachter, & Spradlin, 1988; Spradlin & Saunders, 1986). A study with adult subjects, in which five-member equivalence classes were established in only 2 of 12 cases following linear-series training, also corroborates this prediction (Fields et al., 1995).

Results of other studies, however, appear to be inconsistent with our prediction. For example, Spradlin, Saunders, and Saunders (1992) reported successful development of two five-member classes in normally capable children immediately following linear-series training. Kennedy (1991) employed a "branching" linear-series structure, or what might be considered a combination of linear-series and sample-as-node training, to establish three seven-member classes with normally capable adults. Gradual emergence of equivalence was reported for most subjects, with multinode relations emerging last. Similar results were reported for typically developing children following sample-as-node training (K. J. Saunders et al., 1993). As we noted above and elsewhere (R. R. Saunders & Green, 1992), simple discriminations that were not presented in training may be acquired over the course of testing, even in the absence of differential trial-by-trial consequences. When there are a large number of these, acquisition may be more likely to occur in normally capable adults and older children than in individuals with severe learning difficulties or young children. This could account for the results reported by Kennedy (1991), K. J. Saunders et al. (1993), and Spradlin et al. (1992).

Gradual Emergence of Equivalence

Next, we pick up the thread of our earlier discussion about gradual emergence of equivalence, this time in the context of linear-series training structures. As Table 5 shows, linear-series training (e.g., AB, BC, CD, DE) and testing for two five-member equivalence classes yield 24 simple discriminations that are not presented in training but that are called for on test trials (e.g., A1 vs. C1, B2 vs. D2, C1 vs. E2, etc.). In the terminology of the nodal distance literature, the A versus C, B versus D, and C versus E discriminations are

presented in the one-node tests, the A versus D and B versus E discriminations in the two-node tests, and the A versus E discriminations in the three-node tests. A typical test session intersperses test trials (AC, BD, CA, EC, etc.) among baseline trials (AB, BC, CD, DE) with each trial type appearing equally often. This means that the B, C, and D stimuli will appear three times more often than either A stimulus and 50% more often than either E stimulus on baseline trials during each test session. If previously untrained discriminations develop over the course of testing, as we speculated previously, then we hypothesize that the order in which those untrained discriminations are acquired will correspond to the frequency of reexposure to particular stimuli on baseline trials during testing. That is, one-node tests (BD and DB) should produce positive results first. Other tests involving B, C, and D stimuli (i.e., one-node tests for AC, CA, CE, and EC and two-node tests for AD, DA, BE, and EB) should produce positive results next, and tests involving the A and E stimuli (i.e., the three-node AE and EA tests) should be the last to produce positive results. In other words, patterns of gradual emergence that have been attributed to nodal or associative distance (e.g., Fields et al., 1990; Kennedy, 1991; Kennedy et al., 1994) may instead reflect gradual acquisition of simple discriminations as a function of frequency of stimulus presentation during testing.

To evaluate this possibility, we looked for studies reporting nodal distance effects that conformed to the procedural assumptions underlying our analysis (simultaneous MTS procedures, all baseline trials mixed before testing, test-trial types for each property of equivalence presented together within a session, balanced MTS trials in training and testing, and no differential consequences on test trials). Unfortunately, we found none. The seminal experiment on nodal distance (Fields et al., 1990), for example, employed unbalanced test trials on which comparisons from different stimulus sets were mixed (e.g., CA test trials had A and C stimuli as comparisons). Kennedy (1991) also employed unbalanced trial types. Other studies purporting to show nodal distance effects had other methodological features that made them unsuitable for our discrimination analysis. For instance, extensive pretesting and the use of

stimuli (words) with which subjects had preexperimental histories likely confounded the results (Kennedy et al., 1994). Thus, new experiments will be necessary to evaluate our hypothesis about gradual emergence.

Differential Responding

Exposure to test trials that involve novel discriminations may give rise to behavior that was not the direct product of training contingencies, but that may nonetheless foster acquisition of the new discriminations. For some subjects, preexperimental repertoires may emerge in the presence of novel trial types or arrangements, generating differential responding. Among verbally sophisticated humans, stimulus naming is a common form of such behavior. Although perhaps not necessary for untrained simple discriminations to emerge, differential responding such as naming could foster the acquisition of those discriminations and, therefore, the emergence of equivalence-consistent performances over the course of testing (see Dugdale & Lowe, 1990; Eikeseth & Smith, 1992; Lowe & Beasty, 1987; McIlvane & Dube, 1996; Sidman, 1994).

McIlvane and Dube (1996) suggested that naming or other differential responding may occur even when experimental procedures do not explicitly require it. Several procedures may promote naming. For instance, K. J. Saunders et al. (1993) noted that in prior studies comparing sample-as-node with comparison-as-node training structures (R. R. Saunders, Wachter, & Spradlin, 1988; Spradlin & Saunders, 1986), subjects with mental retardation were given unique names for each stimulus in four "instructed" trials at the very beginning of conditional discrimination training. On those trials, the experimenter named each of the stimuli twice, but did not do so thereafter, and the subjects were never required to repeat the names. Positive outcomes on equivalence tests were seen for all subjects given comparison-as-node training but for only 1 subject given sample-as-node training. In a partial replication, K. J. Saunders et al. (1993) exposed 11 subjects with mental retardation to comparison-as-node training to establish two four-member equivalence classes. Six subjects received four initial instructed trials with stimulus names provided by the experimenter, and 5 did not.

Positive outcomes on equivalence tests were seen for 5 of the 6 instructed subjects but for only 1 of 5 uninstructed subjects. A follow-up experiment showed that when previously unsuccessful uninstructed subjects were given instructed training with new stimuli, they too passed equivalence tests. These results suggest that differential naming of stimuli by the experimenter fostered the establishment of equivalence classes in subjects with mental retardation, including those who received comparison-as-node training (K. J. Saunders et al., 1993). Demonstrations that equivalence classes emerged more readily following sample-as-node training with auditory samples than with visual samples may reflect similar processes (Green, 1990; Sidman et al., 1986). That is, presenting auditory samples (names) in training may set the occasion for subjects to produce names for the other experimental stimuli, thereby fostering simple discriminations among them. These possibilities warrant direct empirical testing.

In a related recent study, R. R. Saunders et al. (1999) provided preschool children with "instructed" training like that provided in previous experiments in the same laboratory (i.e., differential experimenter-provided oral names for each stimulus on the first four training trials). Some subjects had comparison-as-node training, and others had sample-as-node training. The former required more trials for performance to reach the training criterion than did the latter. Similar differences were reported by R. R. Saunders, Wachter, and Spradlin (1988) and Fields et al. (in press). Thus, across comparable experiments, preschool children, normal adults, and adolescents with mental retardation acquired baseline conditional discriminations more rapidly in sample-as-node training than in comparison-as-node training, but were less likely to produce positive outcomes on equivalence tests immediately after training. On the other hand, as noted above, equivalence classes were not established in all the subjects with mental retardation who had comparison-as-node training without instructions that included differential names for the stimuli in the study by K. J. Saunders et al. (1993). Taken together, these results suggested to R. R. Saunders et al. (1999) that instructions may enhance training-structure differences. Because comparison-as-node training presents a

larger number of successive (sample) discriminations than does sample-as-node training, it may be more likely to set the occasion for differential sample naming (at least by verbally skilled subjects). This might foster the development of simple discriminations, not just among the samples but also among samples and comparisons presented simultaneously, which is then likely to produce positive outcomes on tests for the properties of equivalence (cf. Sidman, 1994, pp. 413–414). In contrast, sample-as-node conditional discriminations require fewer successive discriminations than does comparison-as-node training, which might decrease the likelihood that subjects who are capable of naming stimuli will do so during training. This may in turn make them ill-prepared to perform the new discriminations called for on tests. Either or both of these possibilities may account for training-structure differences reported in some stimulus equivalence studies (R. R. Saunders et al., 1999).

Response Speed and Verbal Self-Reports

A recent study used linear-series training (AB, BC, CD, DE, EF, FG) in an attempt to establish three seven-member equivalence classes with 12 college students (Spencer & Chase, 1996). The investigators reported that accuracy of responding on tests for the properties of equivalence decreased with increasing nodal distance for most subjects, but the effect was small and transient. In contrast, speed of correct responding on tests of transitivity and combined tests of symmetry and transitivity was inversely related to nodal distance for nearly all subjects, a relation that was maintained with repeated testing. Subjects responded considerably faster on baseline and symmetry trials than on transitivity and combined trials during test sessions, and somewhat faster on baseline trials than on symmetry test trials. There was no overall difference in speed of responding to transitivity and combined trials.

It is interesting to speculate whether the differential response speeds reported by Spencer and Chase (1996) might have been a function of differential acquisition during training of the simple discriminations required on the tests. The GF symmetry tests, for example, required successive discriminations among G stimuli, which had only been

presented simultaneously (as comparisons) during training. The A, B, C, D, E, and F stimuli, on the other hand, were all presented successively (as samples) during training, so symmetry test trials with those stimuli likely posed little difficulty for the subjects. Because the GF symmetry tests constituted one sixth of all symmetry test trials in the study, decreases in speed on those trial types alone could account for the slight overall differences in speed of responding between baseline and symmetry trials. Indeed, it seems plausible that the decrease in response speed on equivalence test trials that Spencer and Chase attributed to nodal distance was instead a function of the numbers and types of new, untrained discriminations presented on the various test trials. Further, it is possible that some of the simple discriminations that were not explicitly presented in the conditional discriminations trained early in the linear series were nonetheless acquired over the course of subsequent training, whereas those trained late in the series were not. The order in which conditional discriminations were trained was AB, BC, CD, DE, EF, FG. As we suggested earlier, simple discriminations among some stimuli that were never presented together as samples within a session or as comparisons within trials (e.g., the B and D stimuli) might develop when those stimuli are presented successively across trials (e.g., in sessions mixing AB, BC, and CD training trials, as in the Spencer and Chase study). This possibility could not arise, however, for stimuli introduced late in the linear series (e.g., F and G stimuli). Our analysis would predict, therefore, that subjects might respond faster on two-node tests involving stimuli introduced early in the series (e.g., BD, DB) than on two-node tests involving stimuli introduced later (e.g., DG, GD), because the former involve simple discriminations that were more likely to have been acquired than the latter. Response speed on the trials with stimuli from the middle of the training series (BE, EB, CF, FC) should fall in between. All three-, four-, and five-node tests would involve some of the stimuli introduced late in the training series, so they should be excluded from such an analysis.

The differential response speeds reported by Spencer and Chase (1996) seem to parallel verbal self-reports of accuracy on symme-

try and equivalence tests demonstrated in another study. College students who were given sample-as-node training and testing to potentially establish two four-member equivalence classes were asked to report verbally whether they believed their test-trial responses were correct or incorrect (Lane & Critchfield, 1996). Symmetry and equivalence test results (i.e., MTS performances) were highly accurate. Self-reported evaluations of accuracy ranged from 96% to 100% on symmetry tests and 79% to 100% on equivalence tests. That is, despite nonverbal responses on equivalence test trials that were highly consistent with equivalence, verbal reports seemed to reflect some uncertainty about the accuracy of those responses, more so than on symmetry tests. Lane and Critchfield's results are difficult to interpret, however, because of the composition of the training trials. Although training was designed to establish only two classes, three-choice MTS procedures were employed instead of balanced two-choice procedures. This meant that stimuli from different stimulus sets were mixed as comparisons: AB trials had the C stimuli as the third comparisons, AC trials included D comparisons, and AD trials had B comparisons. Thus, unlike sample-as-node training with standard procedures in which the B, C, and D stimuli are never juxtaposed within trials, Lane and Critchfield's training procedures might have established simple discriminations among all the experimental stimuli prior to testing. Further research will be necessary to determine how such unbalanced trial configurations might influence the numbers and types of simple discriminations established in training.

SUMMARY AND CONCLUSION

We have suggested, based on an analysis of the simple discriminations that make up the trained and tested conditional discriminations in typical stimulus equivalence experiments, that sample-as-node and linear-series training are less likely to produce positive outcomes on all tests for the properties of equivalence than is comparison-as-node training. These differential outcomes should be more likely or more marked when training is designed to establish relatively large classes or

a large number of classes. The discrimination analysis also suggests an alternative interpretation of some findings in the stimulus equivalence literature that have been attributed to such variables as nodal distance. For example, we have suggested that differential response speeds or latencies on various types of test trials may reflect differential acquisition of the component simple discriminations that are due to training structure. Unlike nodal distance, this is a characteristic of experimental procedures that relates to familiar behavioral principles.

Further, some specific, testable predictions follow from our main hypothesis:

1. Sample-as-node training should yield less accurate performances on tests for symmetry than does linear-series training, provided class size and number of classes are equated.

2. Following sample-as-node training, symmetry tests present successive discriminations among former comparison stimuli serving as samples, and simultaneous discriminations among former samples serving as comparisons. Positive outcomes on these symmetry tests entail demonstration of the simple successive and simultaneous discriminations necessary for positive results on equivalence tests. Thus, following sample-as-node training, if subjects are given equivalence tests only after producing positive outcomes on the symmetry tests, all test outcomes are likely to be positive. If equivalence tests are given first, however, positive outcomes are less likely.

3. Any procedure that leads to differential responding to each of the stimuli in the experiment, whether explicitly arranged by the experimenter or arising from subjects' preexperimental histories, should establish simple discriminations among all the stimuli, thereby mitigating the differential effects of training structures.

Although our review suggests that published research on stimulus equivalence provides some support for the discrimination analysis of training-structure effects, the evidence is neither overwhelmingly confirmatory nor disconfirmatory. This may reflect the complexities of equivalence research at least as much as the validity of our analysis. In this now sizable body of research, procedural and subject variability is so great that it is very difficult to make comparisons across studies or to draw general conclusions. For

example, we suggested that experiments designed to produce small numbers of small equivalence classes should produce positive outcomes regardless of training structure, because of their relatively low discrimination demands in terms of both the total number and the types of component simple discriminations required. Yet Fields and his colleagues have consistently reported negative outcomes from such experiments, even though their subjects were normally capable adults who should not have had difficulty learning the requisite discriminations (e.g., Fields et al., 1992). One striking difference between those experiments and many others in the basic stimulus equivalence literature is that the stimuli were printed trigrams rather than arbitrary forms or sounds. This suggests that specific characteristics of the stimuli might account for the high rate of equivalence test failures in the Fields laboratory. In particular, the presence of some identical or physically similar letters among trigrams might make for especially difficult simple discriminations among experimental stimuli that we contend are necessary for positive equivalence outcomes. We have not reanalyzed the experiments from the Fields laboratory for this possibility, but our analysis suggests it may be a plausible explanation for the reported equivalence failures (and see K. J. Saunders et al., 1993).

Another study that appears to contradict our analysis was reported recently by Arntzen and Holth (1997). Their data are consistent with our contention that linear-series training is not very likely to produce positive equivalence outcomes, but are inconsistent with our predictions about the outcomes of sample-as-node and comparison-as-node training structures. Following training to establish two three-member classes, Arntzen and Holth reported positive test outcomes in fewer subjects following comparison-as-node training than sample-as-node training, and in even fewer following linear-series training. As R. R. Saunders et al. (1999) noted, however, these investigators conducted their tests in isolation, that is, in blocks of test trials rather than with test trials interspersed among training trials. This raises a question as to how well the trained conditional relations were maintained during testing. Given the observation that comparison-as-node training arranges

more difficult discriminations (i.e., more successive discriminations among samples) than sample-as-node training, it seems plausible that the stability of the baseline performances engendered by the two training structures differed at the point at which the isolated test trials were presented. This might account for the smaller number of subjects in whom equivalence classes were established following comparison-as-node training than following sample-as-node training.

A more recent study by the same investigators poses some interesting challenges to our discrimination analysis. Holth and Arntzen (1998) reported that different mixtures of arbitrary stimuli and familiar stimuli (e.g., pictures of common objects) in linear-series training structures (AB, BC) produced different results with normally capable adult subjects. Recall that this structure yields test trials for transitivity (AC) and equivalence (CA) that require both simultaneous and successive discriminations among the A and C stimuli not presented in training (see Table 1 and Figure 4). When familiar pictures constituted the A and C stimulus sets, the B set, or all three sets, positive equivalence test outcomes were produced more often than when only the A or the C stimuli were familiar pictures, or when all the stimuli were arbitrary. Enhanced outcomes with familiar pictures in both the A and C positions or in all three positions in the linear series are entirely consistent with our analysis: Subjects presumably could discriminate among all those stimuli before the experiment began, so they should have had no difficulty making the simple discriminations called for on the AC and CA tests. When just the A stimuli or the C stimuli were familiar to subjects prior to the experiment, there may not have been enough pre-existing simple discriminations to carry them through the AC and CA tests. It is not readily apparent to us, however, why equivalence test outcomes were also enhanced when just the B stimuli, but not the A and C stimuli, were familiar to the subjects (i.e., discriminated) when they entered the experiment, because the B stimuli did not appear at all on the tests for transitivity and equivalence. Whatever the explanation, the Holth and Arntzen experiment suggests an interesting approach to examining simple discrimination effects within various training structures.

We hope the analysis presented here will prompt research to resolve some of the apparent inconsistencies in the stimulus equivalence literature, particularly inconsistencies that appear to be the bases for ongoing debates about such issues as naming, gradual emergence, nodal or associative distance, and of course, training-structure effects. We do not mean to suggest that all of the reported differences in stimulus equivalence outcomes can be accounted for by differences in the simple discriminations presented in the various training structures; other potential sources of extraneous stimulus control should be examined as well. We do mean to suggest that the role of simple discriminations as inherent components of conditional discriminations has been underappreciated, and that careful attention to this factor can lead to improved experimental designs for stimulus equivalence work.

REFERENCES

- Adams, B. J., Fields, L., & Verhave, T. (1993). Formation of generalized equivalence classes. *The Psychological Record, 43*, 553–566.
- Annett, J. M., & Leslie, J. C. (1995). Stimulus equivalence classes involving olfactory stimuli. *The Psychological Record, 45*, 439–450.
- Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record, 47*, 309–320.
- Barnes, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record, 44*, 91–124.
- Barnes, D., Browne, M., Smeets, P., & 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., McCullagh, P. D., & Keenan, M. (1990). Equivalence class formation in non-hearing impaired children and hearing impaired children. *The Analysis of Verbal Behavior, 8*, 19–30.
- Brady, N. C., & Saunders, K. J. (1991). Considerations in effective teaching of object-to-symbol matching. *Augmentative and Alternative Communication, 7*, 112–116.
- Buffington, D. M., Fields, L., & Adams, B. J. (1997). Enhancing equivalence class formation by pretraining of other equivalence classes. *The Psychological Record, 47*, 69–96.
- Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 58*, 183–204.
- Carter, D. E., & Eckerman, D. A. (1975). Symbolic matching by pigeons: Rate of learning complex discriminations predicted from simple discriminations. *Science, 187*, 662–664.
- Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior, 46*, 243–257.
- Dube, W. V., McIlvane, W. J., & Green, G. (1992). An analysis of generalized identity matching-to-sample test procedures. *The Psychological Record, 42*, 17–28.
- Dugdale, N., & Lowe, C. F. (1990). Naming and stimulus equivalence. In D. E. Blackman & H. Lejeune (Eds.), *Behavior analysis in theory and practice: Contributions and controversies* (pp. 115–138). Hillsdale, NJ: Erlbaum.
- Eikeseth, S., & Smith, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: The role of naming. *Journal of the Experimental Analysis of Behavior, 58*, 123–133.
- Fields, L., Adams, B. J., Newman, S., & Verhave, T. (1992). Interactions among emergent relations during equivalence class formation. *The Quarterly Journal of Experimental Psychology, 45B*, 125–138.
- Fields, L., Adams, B. J., & Verhave, T. (1993). The effects of equivalence class structure on test performances. *The Psychological Record, 43*, 697–712.
- Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior, 3*, 345–358.
- Fields, L., Hobbie, S. A., Adams, B. J., & Reeve, K. F. (in press). Effects of training directionality and class size on equivalence class formation by adults. *The Psychological Record*.
- Fields, L., Landon-Jimenez, D. V., Buffington, D. M., & Adams, B. J. (1995). Maintained nodal distance effects in equivalence classes. *Journal of the Experimental Analysis of Behavior, 64*, 129–145.
- Fields, L., Reeve, K. F., Rosen, D., Varelas, A., Adams, B. J., Belanich, J., & Hobbie, S. A. (1997). Using the simultaneous protocol to study equivalence class formation: The facilitating effects of nodal number and size of previously established equivalence classes. *Journal of the Experimental Analysis of Behavior, 67*, 367–389.
- Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48*, 317–332.
- Fields, L., Verhave, T., & Fath, S. (1984). Stimulus equivalence and transitive associations: A methodological analysis. *Journal of the Experimental Analysis of Behavior, 42*, 143–157.
- Green, G. (1990). Differences in development of visual and auditory-visual equivalence relations. *American Journal on Mental Retardation, 95*, 260–270.
- Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229–262). New York: Plenum Press.
- Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record, 48*, 81–110.
- Innis, A., Lane, S. D., Miller, E. R., & Critchfield, T. S. (1998). Stimulus equivalence: Effects of a default-response option on emergence of untrained stimulus relations. *Journal of the Experimental Analysis of Behavior, 70*, 87–102.
- Johnson, C., & Sidman, M. (1993). Conditional discrimination on equivalence relations: Control by negative

- stimuli. *Journal of the Experimental Analysis of Behavior*, 59, 333–347.
- Kelly, S., Green, G., & Sidman, M. (1998). Visual identity matching and auditory-visual matching: A procedural note. *Journal of Applied Behavior Analysis*, 31, 237–243.
- Kennedy, C. H. (1991). Equivalence class formation influenced by the number of nodes separating stimuli. *Behavioural Processes*, 24, 219–245.
- Kennedy, C. H., Ikonen, T., & Lindquist, K. (1994). Nodality effects during equivalence class formation: An extension to sight-word reading and concept development. *Journal of Applied Behavior Analysis*, 27, 673–683.
- Lane, S. D., & Critchfield, T. S. (1996). Verbal self-reports of emergent relations in a stimulus equivalence procedure. *Journal of the Experimental Analysis of Behavior*, 65, 355–374.
- Lazar, R. M., Davis-Lang, D., & Sanchez, L. (1984). The formation of visual stimulus equivalences in children. *Journal of the Experimental Analysis of Behavior*, 41, 251–266.
- Lowe, C. F., & Beasty, A. (1987). Language and the emergence of equivalence relations: A developmental study. *Bulletin of the British Psychological Society*, 40, A42.
- McIlvane, W. J., & Dube, W. V. (1992). Stimulus control shaping and stimulus control topographies. *Behavior Analysis*, 15, 89–94.
- McIlvane, W. J., & Dube, W. V. (1996). Naming as a facilitator of discrimination. *Journal of the Experimental Analysis of Behavior*, 65, 267–272.
- Michael, R. L., & Bernstein, D. J. (1991). Transient effects of acquisition history on generalization in a matching-to-sample task. *Journal of the Experimental Analysis of Behavior*, 56, 155–166.
- O'Leary, C. A., & Bush, K. M. (1996). Stimulus equivalence in the tactile modality. *The Psychological Record*, 46, 509–517.
- Pilgrim, C., Chambers, L., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: II. Children. *Journal of the Experimental Analysis of Behavior*, 63, 239–254.
- Pilgrim, C., & Galizio, M. (1990). Relations between baseline contingencies and equivalence probe performances. *Journal of the Experimental Analysis of Behavior*, 54, 213–224.
- Pilgrim, C., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior*, 63, 225–238.
- Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record*, 43, 725–744.
- Saunders, K. J., & Spradlin, J. E. (1989). Conditional discrimination in mentally retarded adults: The effect of training the component simple discriminations. *Journal of the Experimental Analysis of Behavior*, 52, 1–12.
- Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: Programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior*, 60, 571–585.
- Saunders, R. R., Drake, K. M., & Spradlin, J. E. (1999). Equivalence class establishment, expansion, and modification in preschool children. *Journal of the Experimental Analysis of Behavior*, 71, 195–214.
- Saunders, R. R., & Green, G. (1992). The nonequivalence of behavioral and mathematical equivalence. *Journal of the Experimental Analysis of Behavior*, 57, 227–241.
- Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Behavior*, 50, 145–162.
- Saunders, R. R., Wachter, J. A., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior*, 49, 95–115.
- Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, 14, 5–13.
- Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.
- Sidman, M. (1987). Two choices are not enough. *Behavior Analysis*, 22, 11–18.
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative.
- Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior*, 43, 21–42.
- Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior*, 37, 23–44.
- Sidman, M., & Tailby, W. (1982). Conditional discriminations vs. matching-to-sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37, 5–22.
- Sidman, M., Willson-Morris, M., & Kirk, B. (1986). Matching-to-sample procedures and the development of equivalence relations: The role of naming. *Analysis and Intervention in Developmental Disabilities*, 6, 1–19.
- Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, 65, 643–659.
- Spradlin, J. E., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal of Mental Deficiency*, 77, 556–566.
- Spradlin, J. E., Saunders, K. J., & Saunders, R. R. (1992). The stability of equivalence classes. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 29–42). Reno, NV: Context Press.
- Spradlin, J. E., & Saunders, R. R. (1986). The development of stimulus classes using match-to-sample procedures: Sample classification versus comparison classification. *Analysis and Intervention in Developmental Disabilities*, 6, 41–58.
- Tierney, K. J., de Lary, P., & Bracken, M. (1995). Formation of equivalence classes incorporating haptic stimuli. *The Psychological Record*, 45, 431–437.
- Urcuioli, P. J., & Zentall, T. R. (1993). A test of comparison-stimulus substitutability following one-to-many matching by pigeons. *The Psychological Record*, 43, 745–759.
- Urcuioli, P. J., Zentall, T. R., Jackson-Smith, P., & Steirn, J. N. (1989). Evidence for common coding in many-

- to-one matching: Retention, intertrial interference, and transfer. *Journal of Experimental Psychology: Animal Behavior Processes*, 15, 264–273.
- Wetherby, B., Karlan, G. R., & Spradlin, J. E. (1983). The development of derived stimulus relations through training in arbitrary-matching sequences. *Journal of the Experimental Analysis of Behavior*, 40, 69–78.
- Williams, D. C., Saunders, K. J., Saunders, R. R., & Spradlin, J. E. (1995). Unreinforced conditional selection within three-choice conditional discriminations. *The Psychological Record*, 45, 613–627.

Received January 23, 1998
Final acceptance March 29, 1999

ERRATUM

Weiss, S. J., & Panlilio, L. V. (1999). Blocking a selective association in pigeons. *JEAB*, 71, 13–24. On p. 15, the frequency of the tone is given as 400 Hz when it should be 440 Hz.