The Analysis of Verbal Behavior 2009, **25,** 5–17

Equivalence Relations and Behavior: An Introductory Tutorial

Murray Sidman, Sarasota, Florida

With an emphasis on procedural fundamentals, the original behavior-analytic equivalence experiments and the equivalence paradigm are described briefly. A few of the subsequent developments and implications are noted, with special reference to the possible significance of the findings with respect to language and cognition.

Key words: equivalence relations, stimulus control, conditional discrimination, matching to sample

I was asked to do a brief introduction to equivalence relations in behavior and its implications for language research and application, particularly for the benefit of young readers of this journal who are becoming more interested in the topic. This introduction to research on equivalence relations is therefore going to be extremely basic, starting at the very beginning and emphasizing methodology. Expectations of a big theoretical discussion will only result in disappointed readers. There is still much to be done in equivalence research and in its applications that is independent of any particular theory. We are, of course, interested in what any data signify, but there are many kinds of significance besides theoretical. I will note some exciting possibilities that I see in the topic of equivalence relations with respect both to the science of behavior and to more general intellectual and practical concerns. What remains to be done is at least as stimulating as what has been done already. Even after an exciting research program that has now lasted more than 35 years, I am eager to see others expand on the basics.

For those who want to follow up in more detail, two references that give my own slant on the field of equivalence relations are my equivalence book (Sidman, 1994) and a paper that expands on some of the material in the book (Sidman, 2000). I will start here with a description of our first experiment (Sidman, 1971). Although many investigators have since done more sophisticated and more revealing studies, the first ones have certain virtues as an introduction. Even the very first had features that are still relevant to what is being done today. For those who want to go more deeply into the initial data, a more fully controlled replication was published 2 years later (Sidman & Cresson, 1973).

Our basic procedure was matching to sample. That term, *matching to sample*, referred originally to what experimenters thought of as identity matching, in which subjects have to match stimuli that, to us, are physically the same. In most of our experiments, although not all, the stimuli to be matched bore no physical resemblance to each other. Because the matching criteria were arbitrary, I prefer the procedural name, *conditional discrimination*. If you are given Stimulus A1, then you match it to Stimulus B1 and not to B2, B3, or B4. If you are given Stimulus A2, however, then you match it to B2 and not any of the others. If A3, then B3, and so on. The experimenter or teacher determines which stimuli are to be related, and the matching is done regardless of any lack of resemblance between the matched stimuli.

The involvement of arbitrary matching brings up what many consider to be the most interesting aspect of equivalence relations. The emergence of equivalence relations provides a way to study experimentally what might be thought of as a kind of stimulus generalization, an elusive kind in which subjects come to match stimuli that share no physical properties and that have never been paired with or directly related to each other.

Here is an outline of the experimental setup we started with. On any given conditional discrimination trial, the subject was to compare several stimuli (called comparison stimuli) to a sample and to select one of those comparison stimuli by touching it; the choice

Address correspondence to Murray Sidman, 3435 Fox Run Road #347, Sarasota, Florida 34231 (e-mail: murraysidman@comcast.net).

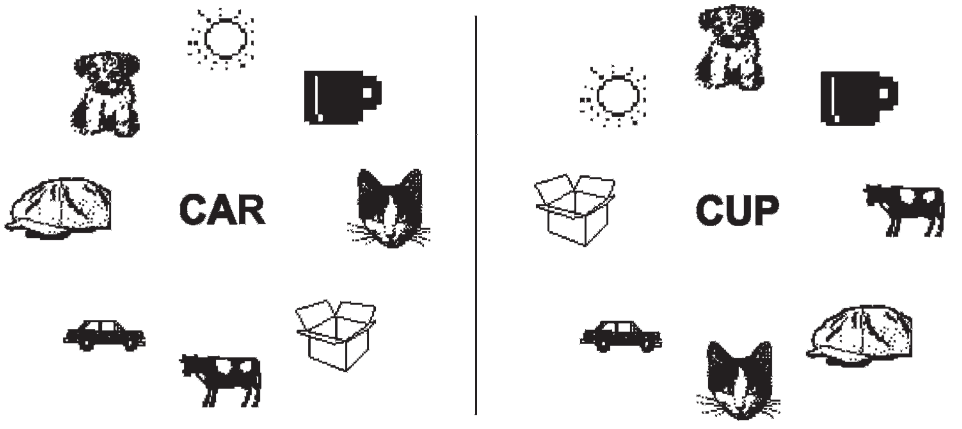

Figure 1. Stimulus displays from two trials of visual-visual word-to-picture matching. The center key on the left contains the sample word *car*, and on the right, the sample word *cup*. Eight comparison stimuli surround each sample.

that we scored correct was conditional on which stimulus was the sample on that particular occasion. In our laboratory, subjects sat before a matrix of nine keys onto which stimuli could be projected, one in the center surrounded by eight others. As an example, Figure 1 shows two trials. At the left, with the sample word CAR located in the center of a circle of eight keys, the subject is to select the picture of a car from among the comparisons located in the outer keys. At the right, with the sample word *cup*, the picture of a cup is to be selected. On other trials, with different samples (*cat, box, cow*, etc.), different selections from among those same comparisons will be correct. The locations of the comparison pictures change from trial to trial. Because the conditional discrimination terminology is somewhat cumbersome, we often still talk about matching to sample even though we are studying nonidentity matching.

Skinner (1950, pp. 213–214) found that he could not easily get pigeons to do matching to sample unless he taught them first to peck the sample key to gain access to comparison stimuli. This procedure was probably effective because pigeons usually look at whatever they peck, so pecking the sample may have helped ensure that they observed the sample. Based on Skinner's finding—along with human subject replications, particularly with normal and handicapped children—today's standard matching-to-sample procedure requires even human subjects to respond to the sample before the comparisons can appear. Requiring human subjects to touch the sample key does not, however, guarantee that they will observe the stimulus on the key. I suspect that such failures of observation are responsible for the seeming inability of some subjects to learn a particular matching-to-sample task—they just do not look at the stimulus on the sample key. Even when they are performing a conditional discrimination perfectly, the stimulus aspects that control their behavior may not be the same as those specified by the experimental contingencies (e.g., Carrigan & Sidman, 1992; Iversen, Sidman, & Carrigan, 1986; Johnson & Sidman, 1993; McIlvane & Dube, 2003). Observing behavior in matching to sample is a ripe area for investigation (Dube et al., 2006). How can it be measured, taught, and modified?

The first phase of our experiment was to check whether our subjects could match printed word samples to picture comparisons, as summarized in Figure 1 and in the left side of Figure 2. This is an instance of what we call visual-visual word-to-picture matching. We also tested another example of nonidentity matching to sample—the reverse, or symmetric version of what we have just been looking at (Figure 2, right). Now, the sample is a picture, and the comparisons are printed words. This is an example of what we call visual-visual picture-to-word matching.

These kinds of stimuli interested us because people who can match printed words to

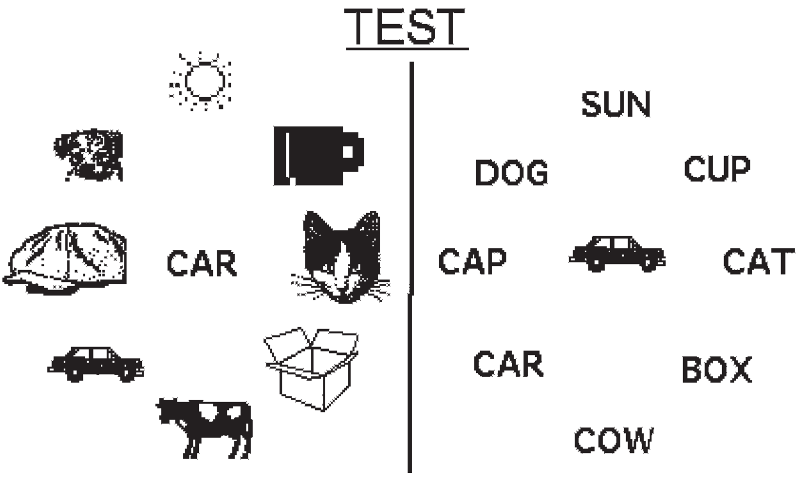

Figure 2. Stimulus displays from pretests of visual-visual word-to-picture matching (left) and visual-visual picture-to-word matching (right).

the appropriate pictures are said to understand the words, to exhibit a simple form of reading comprehension. For reasons that need not concern us here, we wanted to find out whether elementary reading comprehension, tested this way via word-to-picture and picture-to-word matching, could develop without being directly taught. Could we get students to do these matching tasks without our ever having provided them with any reinforcing consequences for doing so?

We had been working with a group of institutionalized teenaged boys with severe mental retardation, boys who were unable to do the two tasks illustrated in Figure 2. They could not match printed words to their corresponding pictures; they had never learned to read. Indeed, before we could get them to do the complex matching to sample that this experiment required, we had to teach them basics like sitting quietly, pointing at specific objects, discriminating simple forms like lines of different orientations and curvatures, and telling circles, squares, and other standard forms from each other. Before we could expect them to discriminate words, we had to teach them to discriminate the individual letters, and before that, the forms that make up the letters. Finally, we had brought them to the point where we could teach them to do what was called identity matching—to match words to themselves and pictures to themselves—so they had become familiar with our matching-to-sample procedures. Because they had shown no evidence of reading comprehension, they seemed ideally suited to help answer our question about how to teach it.

We first taught the boys to match dictated word samples to picture comparisons (Figure 3). Instead of presenting visual samples on the center key, which remained blank, we presented auditory samples, dictated words. On the particular trial shown at the left, we dictated, ''car, car, car, …,'' repeating the word until the end of the trial so that the boy would not have to remember it. On the right side is an example of another trial, this one with the dictated word ''cup'' as the sample. Before the comparison stimuli could appear, the boy had to touch the blank sample key and thereby produce comparison pictures on the outer keys, and we required at least one sounding of the sample word before a touch to the blank key would work. The reason we did that was to decrease the likelihood that a boy would impulsively press the blank sample and a comparison key without having a chance to listen to the spoken word. If he did not listen, he would not have an opportunity to learn anything about the dictated samples.

Then, having produced comparison stimuli on the outer keys, the boy could produce a reinforcer by touching the comparison picture that matched the dictated sample. On trials other than those shown in Figure 3, we dictated other names. The subjects eventually

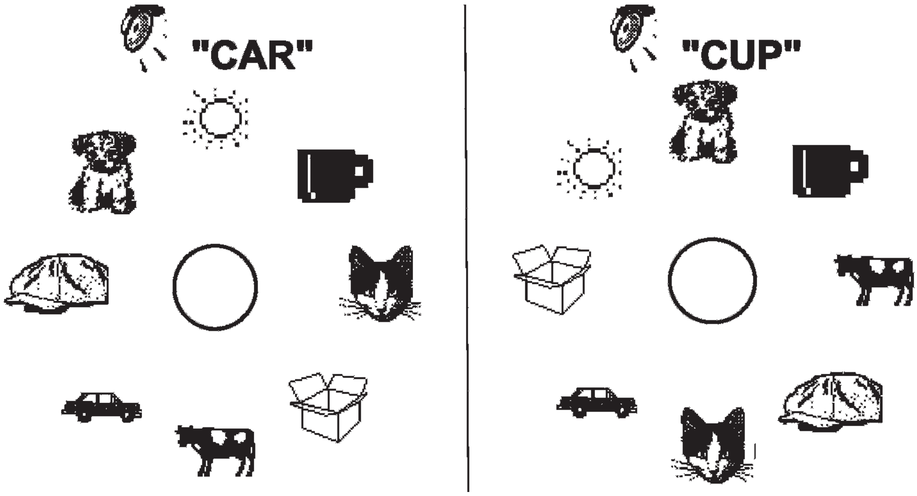

Figure 3. Comparison-stimulus displays from two teaching trials of auditory-visual word-to-picture matching. The center keys are blank and the dictated sample words *car* and *cup* are indicated above the comparison displays (here only, but not to the subjects).

learned to match 20 dictated names to corresponding pictures. We also used several variations of each picture, so that the boy would not just observe some irrelevant aspect of a picture. Let us call this auditory-visual word-to-picture matching.

The next step was to teach them to match the same dictated words not to pictures but to printed words. The right side of Figure 4 shows this. Again, on the illustrated trial, we repeatedly dictated the word ‘‘car.’’ On other trials, we dictated other words. The boy could now procure a reinforcer by touching the corresponding printed word rather than a picture. This task, which we call auditory-visual word-to-word matching, was extremely difficult to teach to our first subjects, but they eventually learned to match the 20

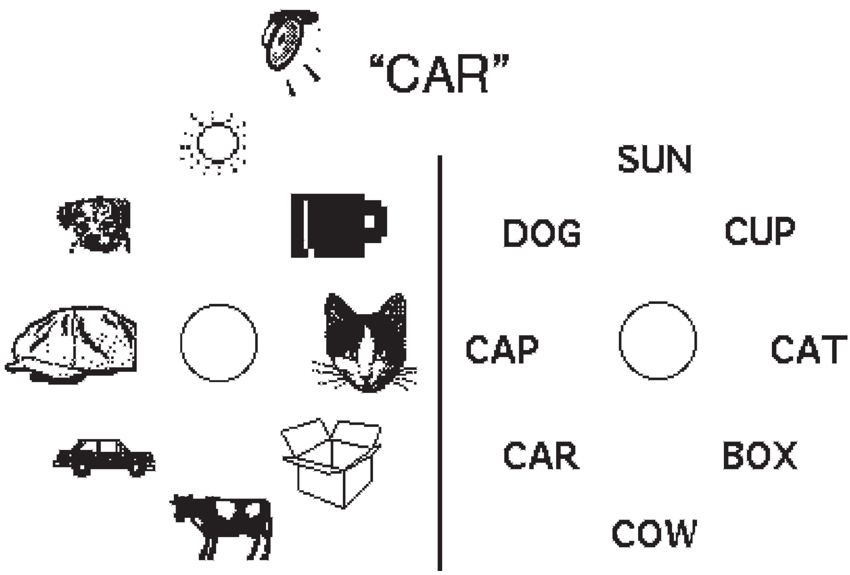

Figure 4. Comparison stimulus displays from teaching trials of auditory-visual word-to-picture matching (left) and auditory-visual word-to-word matching (right).

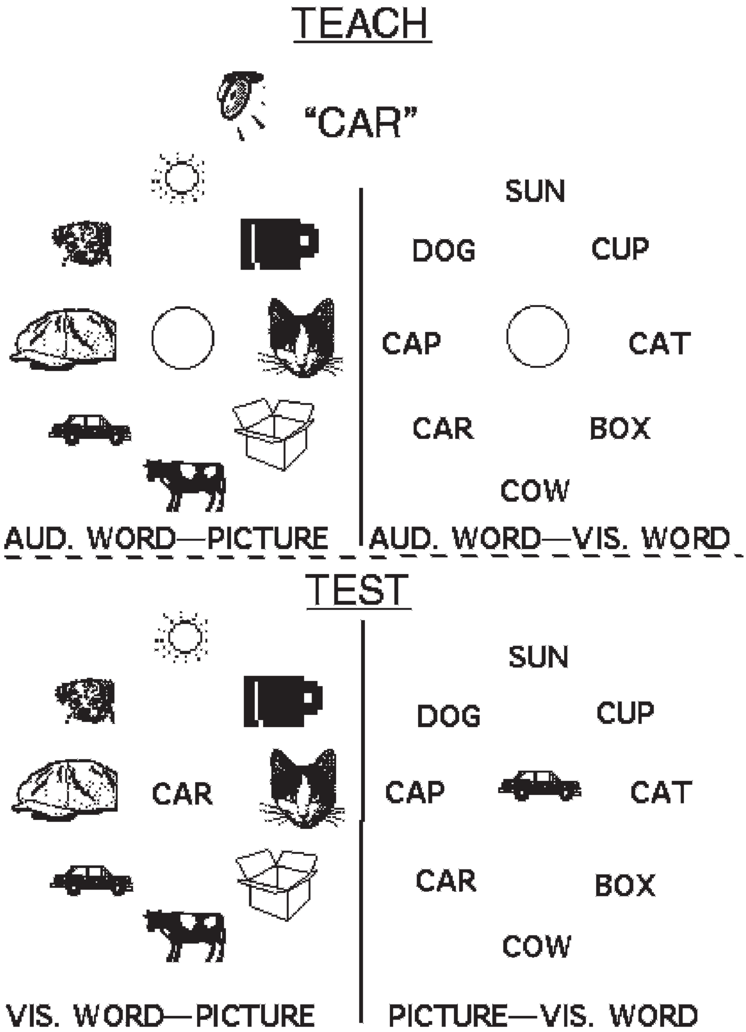

Figure 5. Procedural summary of the first experiments. The two upper segments illustrate the auditory-visual word-to-picture and word-to-word teaching trials that were shown in Figure 4, and the two lower segments illustrate subsequent posttest trials of visual-visual word-to-picture and picture-to-word matching.

dictated names with the corresponding printed names.

At this point, some might be tempted to say that the boys had learned to read text (i.e., printed words). Indeed, we found later that learning these auditory-visual matching tasks often, although not always, did make subjects able to name the printed words (i.e., to read them aloud). But we could not say yet that the boys understood the words, that they were reading with comprehension. For example, I can match many words spoken to me in German with their printed counterparts and I can also read many German words aloud with something resembling the German pronunciation—all this, however, without having the slightest idea what those words mean. We had not yet shown that the boys grasped the relation between printed words and pictures, which would have indicated at least a simple understanding of the words. To find out if they could now read with comprehension, we repeated the original visual-visual word-to-picture and picture-to-word matching tests that we saw earlier in our procedure illustrations. In Figure 5, the upper section repeats the teaching phase. That phase had involved only auditory-visual matching: dictated words to pictures and to printed words. The lower section of Figure 5 illustrates one trial from each subsequent test, which involved only visual stimuli and no auditory samples. Given a printed word as the sample (shown at the lower left, in the center key), would the boy now select the appropriate picture? And given a picture sample (shown in the lower right), would he now select the appropriate printed word? The boys had never been able to do these kinds of visual-to-visual matching tasks before. If they could now do them, we would be able to assert that learning to match dictated words both to pictures and to printed words had given them the ability to comprehend the printed words, to match them to their corresponding pictures.

That is exactly what happened. Although our students had never been taught explicitly to relate text and pictures, they now accurately matched nearly every one of the 20 printed-word samples to its picture and each of the 20 picture samples to its corresponding printed word. After they had learned the original 40 auditory-visual relations via direct teaching with reinforcement, 40 new visual-to-visual relations literally emerged—in full bloom, so to speak. They could now read with comprehension without their doing so ever having been reinforced.

This was not the usual transfer-of-training phenomenon. It was not that the auditory-visual experience permitted the boys to learn the visual-visual matching faster than they otherwise would have. They matched the pictures and printed words perfectly on the very first posttest trials; they showed reading comprehension immediately. Nor was this the usual stimulus generalization phenomenon. It could not be said that the visual-visual relations between printed words and pictures emerged because of any physical resemblances between related stimuli. In everyday language, we could say that the printed words had become symbols for the pictures.

The first time I saw this happen was a big event in my life. For me, it was an experience comparable to the first time I shaped a rat's bar pressing, and then to the first time I conditioned avoidance behavior with the free-operant procedure. My excitement also matched what I felt the first time I taught a difficult circle-ellipse discrimination errorlessly by means of a stimulus-fading procedure, and then, to the first time I found that some patients who had suffered strokes and had lost the ability to express themselves vocally could nevertheless understand words when they were tested nonvocally, that is to say, when they were tested with exactly the same matching-to-sample procedures we have been looking at here. Although they could not say the names, some of them could still match the printed words and pictures.

I am convinced that the best way to get one's feet wet in equivalence, to experience the same excitement I did during our first equivalence experiment, would be to do this experiment oneself. One can do it easily with fewer stimuli and with table-top procedures. Just seeing it happen would be more likely to stimulate interest than would any exposure to published papers, lectures, or theoretical controversies, especially to presentations concerned with advancing some particular theory. I believe that such personal exposure would also generate interest not just in equivalence but, more generally, in behavior analysis. Much remains to be done both in equivalence research and its applications that is independent of any particular theory. An overemphasis on theory has caused our field to overlook a number of interesting published extensions of equivalence relations because the publications require readers to wade through a complex theoretical background in order to find out how the data were related to what was done rather than to their theoretical rationale. For the most part, then, I will just share a number of conjectures about the general significance of the phenomenon that our first experiments revealed.

First, however, a few things need saying about methodology. My concern with methodology arises from observations that the methods we use to gather and present evidence may not only influence scientific conclusions and judgments but may also determine what we do or fail to do next. Take, for example, the equivalence triangle that we often use to summarize a basic equivalence procedure and its findings (Figure 6). Here, A designates three of the dictated word samples we used in the first experiments; B designates three of the pictures, and C

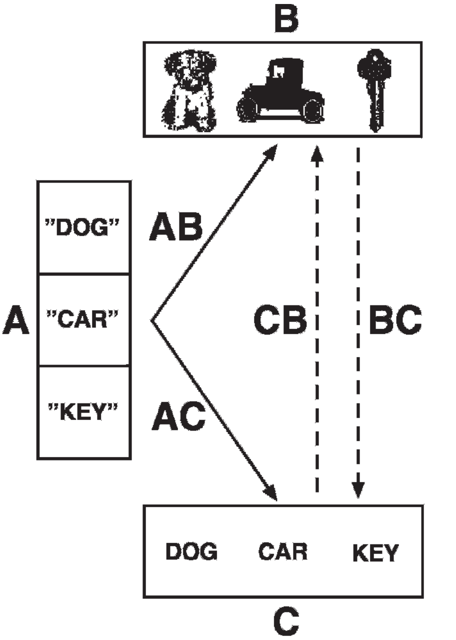

Figure 6. The equivalence triangle. The solid arrows (AB and AC) designate conditional discriminations that were explicitly taught to the subjects: auditory-visual word to picture (AB) and auditory-visual word to word (AC). The dashed arrows indicate conditional discriminations that emerged without having been actually taught: visual-visual word to picture (CB) and visual-visual picture to word (BC).

designates three of the printed words. The two solid arrows indicate the auditory-visual word-to-picture (AB) and word-to-word (AC) matching that were explicitly taught to the subjects. The two dashed arrows indicate the visual-visual word-to-picture (CB) and picture-to-word (BC) matching that emerged without having been explicitly taught.

Unfortunately, the use of arrows in such diagrams suggests to many that equivalence relations represent sequential processes. What is intended, however, is to indicate contingencies, events that are true only under certain conditions: ''If this, then that; if not this, then not that.'' For example, if the defined sample is *dog* and not any of the other possibilities, and if the picture of a dog and not any of the other pictures controls the defined response (touching), then and only then will the defined reinforcer be forthcoming. If *car* and not any other possibility is the sample, then reinforcement becomes subject to a different set of conditions. Summary diagrams like Figure 6 do not show the actual contingencies, and the arrows may easily lead one to ignore the procedural complexities and talk about temporal sequences and associations rather than about simultaneous options. One will then be less likely to ask questions about the contextual control of equivalence relations, about the role of unintended instructional control of the emergence of new conditional discriminations, about the number of possible classes as a determiner of how quickly new conditional discriminations emerge and even as a critical factor in the generation of new conditional discriminations without reinforcement, and about a number of other likely extensions of the basic phenomenon outside the laboratory. In addition, as careful reading of many theoretical discussions will reveal, oversimplification of methodological and procedural descriptions will produce oversimplification of theoretical formulations also.

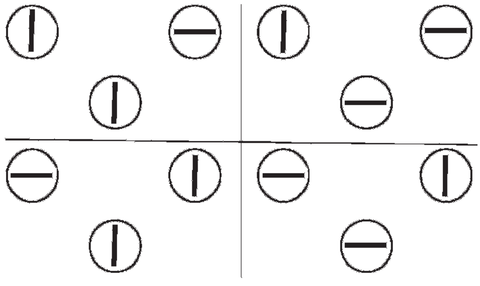

Figure 7. The four possible stimulus displays in the process of teaching horizontal–vertical identity matching.

Methodological considerations not only prove relevant to the evaluation of data in equivalence experiments but sometimes prove interesting in their own right, with extensions also to other research areas. For example, studies of equivalence relations do not always require such complicated stimuli as we have been looking at, or so many stimuli. Such technical simplification can, however, introduce complexities of data interpretation. At one stage of our work, we tried to make matching to sample easier for our subjects by using only two comparison stimuli per trial. For example, Figure 7 illustrates the four possible trial displays from an attempt to teach subjects to match vertical lines to vertical, and horizontal to horizontal, a seemingly simple identity-matching task. The circles represent keys, and a vertical or horizontal line appears on each key. In each trial display, the horizontal and vertical comparison lines are at the top, above the sample; the left and right positions of the comparisons vary from trial to trial. Subjects produce a reinforcer if they touch a comparison key that contains a line with the same orientation as the sample key. What kind of a performance would allow us to say that our subject knows how to match vertical to vertical and horizontal to horizontal?

Suppose he or she achieves 75% correct over enough trials to make that statistic significantly different from chance. We would like to think this average signifies that the subject has matched correctly on 75% of all trials, regardless of which line was the sample. With only two comparisons, however, a score of 75% correct could have been achieved in another way. Suppose that when the sample is vertical (as it is in the two displays on the left side of Figure 7), the subject always selects the vertical comparison; that would yield a score of 100% on vertical-sample trials. But suppose that when the sample is horizontal (as it is in the two displays on the right side of Figure 7), he or she always selects the left comparison, regardless of which stimulus is in that position. Because each comparison appears at the left on half the trials, that would give a score of 50% correct when the sample is horizontal, even though really the subject never selected the horizontal comparison at all; he or she just picked the left comparison key, no matter which line was on that key. With 100% correct on vertical trials and 50% recorded as correct on horizontal trials, the average score would be 75%, even though the subject had never, on any trial, paid any attention to the horizontal comparison. The 75% score, even if statistically significant, is behaviorally meaningless. Whenever I see an author claiming that an accuracy of 75% (even 80%) indicates that a subject has learned a two-sample two-comparison conditional discrimination, I stop reading that paper (for a more detailed discussion, see Sidman, 1980).

With only two comparisons, a related misconception can arise if a subject were always to select the same comparison stimulus. Suppose our student chooses vertical on every trial, regardless of which line is the sample, giving a score of 100% on trials with vertical samples and 0% with horizontal samples. Although this yields an average score of 50%, statistically insignificant, such a performance has been known to tempt investigators into concluding that although the student had completely failed to learn the relation between the horizontal stimuli, he or she had learned to match vertical samples perfectly. A student who always selects the same comparison stimulus, however, cannot be said to be matching either of the samples, horizontal or vertical. Such a performance could indicate simply that the student paid no attention at all to the samples, that as far as he or she was concerned, the samples did not even exist. After all, picking the vertical comparison all the time produced a reinforcer on every other trial on the average; not a bad payoff for so little work.

What if our subject never makes a mistake, always picking the comparison line that matches the sample? Still, with only two comparisons, even a seemingly perfect performance does not permit us to say for sure that he or she has matched both line orientations. Let me describe how I might get a reinforcer on every trial without matching each comparison line to its identical sample. Suppose that on every trial, I look for the vertical comparison. When I find it, I touch it if the sample is also vertical. But if the sample is not vertical, then, whatever else the sample may be, I touch the other comparison, whatever it may be, as long as it is not vertical. Note that all of my selections here are controlled by just one of the stimuli, the vertical line; I either select the vertical comparison or reject it, depending on whether the sample is vertical. If you were then to test me by substituting even unfamiliar stimuli for the horizontal line, I would still be correct on every trial because the only thing that mattered to me was whether or not I was looking at vertical.

The recorded measure, accuracy, does not distinguish between the two types of stimulus control on correct trials: selection of one comparison or rejection of the other. The response that we record, touching the correct comparison, does not tell us whether that comparison or the other one controlled our choice. If you assumed that my perfect score meant I had learned to match vertical and horizontal comparisons to their identical samples, you would be mistaken. If you then tried to build on those stimulus control topographies to teach me something else, you would run into serious problems. For example, the use of only two comparisons per trial might then cause a failure to demonstrate equivalence relations, simply because the actual controlling stimuli are not the ones you are testing for (Carrigan & Sidman, 1992; Johnson & Sidman, 1993). The question, ''What are the actual controlling stimuli?'' remains relevant not just in equivalence research but in every experiment that involves stimulus control (McIlvane & Dube, 2003; Ray & Sidman, 1970). Identifying the actual controlling stimuli is also critical in applied situations, particularly when one is trying to remediate seeming failures to learn.

These are instances in which an aspect of the research methodology (e.g., the use of only two comparison stimuli) can greatly increase the ambiguity of one's conclusions. By presenting three comparisons with each of three samples, one can reduce the likelihood of such a problem. Rather than learning to reject one of two comparisons, the subject must then learn to reject two of the three comparisons on each trial; rejection becomes more difficult than selection. I have, however, seen subjects matching two of the comparisons appropriately to their respective samples, by selection, and then rejecting both of those comparisons on trials with the third sample, thereby learning nothing about the relation between the third sample and comparison. In my own work, therefore, I have gone back to using at least four samples and four comparisons per trial, and I now realize how lucky I was to have started with displays of eight comparisons.

Back to some additional but still elementary matters of significance. The procedures I have outlined have great generality. An extension that to me seems obvious is the teaching of simple vocabularies, an extension whose utility has gone unnoticed within the education establishment. Indeed, very few applied behavior analysts seem to have recognized this particular application. The basic equivalence paradigm provides a most efficient way to add nouns to someone's

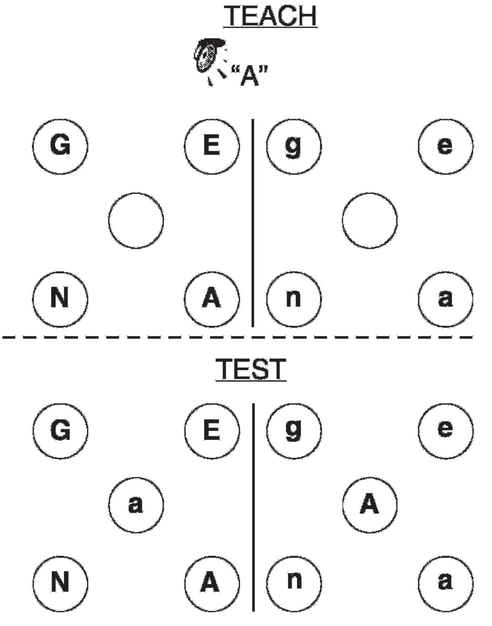

Figure 8. Like Figure 5 but with dictated letter names, upper case letters, and lower case letters as stimuli. After being taught to match dictated letter-name samples to both upper case and lower case comparisons (the two upper segments), posttests revealed the subjects' emergent ability to match the upper and lower case letters to each other (the two lower segments).

reading vocabulary and probably to his or her speaking vocabulary also. Teach them to match spoken words both to their corresponding printed words and pictures, and without any more instruction, they are able to understand the printed words. The same could easily be done with adjectives, adverbs, and by using the capabilities of modern computers, even verbs. Equivalence classes have been generated with many different stimulus materials, not just the strange stimuli that experimenters often use to reduce the likelihood of preexisting equivalence relations. For example, we have taught children with retardation to match both a color and its printed name to the same dictated color name, and have then seen the children able to match colors and printed names to each other, in other words, to understand printed color names.

By teaching children to match dictated letter names to both upper and lower case letters (Figure 8, top), we have then seen

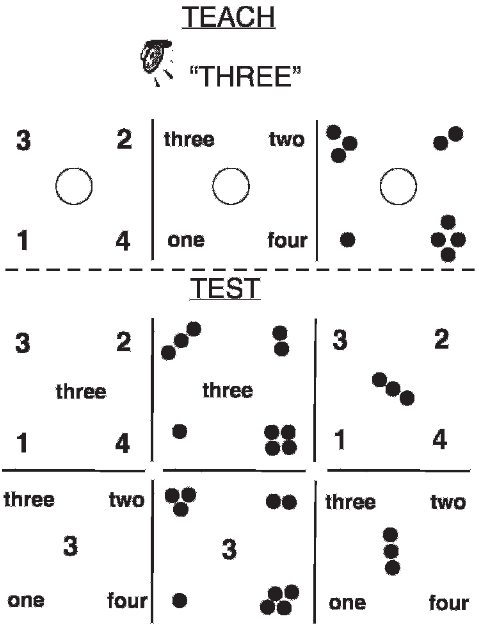

Figure 9. After learning to match dictated number-name samples to comparison digits, printed number names, and quantities (the upper three segments), subjects show in subsequent tests their emergent ability to match each of the former comparisons to each other (the lower six segments).

them able, without further instruction, to match those visual stimuli to each other, as in the lower segment of Figure 8, in other words, to understand the equivalence of upper and lower case letters.

The application of the basic conditional discrimination equivalence procedure to numbers allows us to take a significant step forward. Suppose we first teach students to match dictated number names both to digits and to printed number names, as illustrated in the leftmost and center segments of the upper section of Figure 9. After such teaching, children are able to match both the digits and printed words to each other, as shown in the two leftmost test segments.

Then, as shown in the rightmost teaching segment, suppose we teach them to match the dictated number names also to quantities, represented here by quantities of dots. After this additional teaching step, we find that the classes have enlarged. As indicated in the four right segments of the test section (Figure 9), the children now can also match both the printed digits and number names to the dot quantities, and vice versa, even though they had not been directly taught to do so and had never before even seen those stimuli (visual words, digits, and quantities) together. By learning to match a new stimulus to one member of a class, the children automatically become able to match the new stimulus to all other members of the class. If we were then to go on and teach our pupils to match dictated number names in another language (say, French) to the digits, they would automatically become able also to match the spoken French number names to the printed English number names and to the quantities. This productivity is one of the most significant aspects of equivalence relations. It is not a theory; it is a datum, an exciting datum. As a class enlarges, the direct addition of just one new member to the class produces an enormous increase in the number of indirectly established new relations. A small amount of teaching can yield a tremendous amount of learning.

As far as I am aware, this explosive feature of equivalence relations has not been purposefully exploited in the teaching of second-language vocabularies. Nor has it been used, except in research, for the teaching of money skills like coin equivalences (McDonagh, McIlvane, & Stoddard, 1984; Stoddard, Bradley, & McIlvane, 1987). For example, by directly teaching students to match the dictated ''twenty-five cents'' with one quarter, five nickels, two dimes and a nickel, two dimes and five pennies, three nickels and 10 pennies, and so on, we can then expect that the students, without additional training, will be able to match each of those coin combinations to all the others. To devise such a program would take a lot of planning and preparation, for sure, but what a payoff! Anyone could do it, even if they had never attended any of those confusing symposia on equivalence theory.

The design of our first experiments led some to assume that one must learn auditory-visual relations before showing visual-visual relations like those involved in reading comprehension. It is now clear, however, that new arbitrary visual-visual relations can emerge not just from auditory-visual relations but also from relations that involve only visual or several other stimulus modalities.

The auditory modality is not required. Class union comes about when two sets of conditional discriminations have an element in common. The process is general, regardless of the stimulus modalities. Nevertheless, I venture to suggest that the use of audition and olfaction with nonhumans might prove especially productive. For example, auditory-visual matching in studies with dolphins could provide a key to our understanding of their communication, including both comprehension and production of speech (that is to say, dolphin speech). With the sophisticated techniques that are now available for recording and reproducing sounds, it should be possible to use natural vocalizations as stimuli, relating them either to arbitrary stimuli and responses or to referents that we suspect are the actual ones for those animals.

Species generality of equivalence relations still requires investigation. The original and most subsequent work on equivalence relations involved humans, with emphasis on questions about reading; about language development; about the origin of symbols, classification, and concept formation; about the sources of some seemingly untaught behavior; and about implications for efficiency in teaching. From the beginning, however, a primary concern has been the possibility of finding a role for equivalence phenomena in the behavior of other species. In spite of the failure of the earliest attempts to demonstrate species generality (e.g., Dugdale & Lowe, 2000; Sidman et al., 1982), confirmation of equivalence relations in individual nonhuman subjects has come from Vaughan's (1988) pigeon experiment, in which he pioneered a simple discrimination rather than a conditional discrimination technique for generating equivalence classes. Sidman, Wynne, Maguire, and Barnes (1989) reported a systematic replication of that technique with human subjects. More recently, equivalence relations have been convincingly demonstrated in several sea lion studies in Schusterman's laboratory (e.g., Kastak, Schusterman, & Kastak, 2001). It has been suggested (Sidman, 2008) that the success with sea lions came about at least in part because the subjects, unlike those in other experiments, had previously been taught to do identity matching. Still to be ascertained, however, is even an approximate range of the species that are capable of developing equivalence classes.

Most, if not all, stimuli are members of more than one class. For example, in some instances green is in the class of color, in other instances it is in the class of beginner, and more recently, it is often in the class of environmentally conscious. Stimulus membership in more than one class does not, of course, cause all of those classes to combine into one. Contextual circumstances—the subject of the particular book one is reading, the current topic of conversation, the particular person with whom one is conversing—determine which of several possible classes a stimulus is in at any particular time. In some contexts, classes that contain a member in common will merge into one; class union will take place. In other contexts, classes will simply intersect, remaining separate in spite of a member common to both. In an example provided by Bush, Sidman, and de Rose (1989), "If we are discussing disciplines, Renoir, Constable, and Pollock go together as artists; Twain, Voltaire, and Byron as writers; and Churchill, Kennedy, and De Gaulle as heads of state. If we are discussing nationality, Renoir, Voltaire, and De Gaulle go together as French; Twain, Kennedy, and Pollock as American; and Churchill, Constable, and Byron as British" (p. 31).

Just as the two-term operant reinforcement contingency rarely, if ever, exists in real life (the discriminated operant is therefore the basic unit of behavior analysis) so equivalence relations probably always come under contextual control. Here is another fertile area for investigation, particularly in extending the significance of equivalence classes beyond the laboratory.

When the topic of contextual control of equivalence relations comes up, I often recall the time I emphasized in a talk that an object and its name are not always equivalent, that context determines class membership. As an example, I pointed out that although we swat flies, we do not swat the word *fly*. Steve Hayes then proposed jokingly that he could disprove my theory of equivalence; he printed the word *fly* on a piece of paper and then batted the word with a fly swatter. Nevertheless, the equivalence of particular words to what, in everyday speech, are referred to as their referents is a sufficiently everyday observation to provide a topic even for popular cartoonists. For example, I remember a comic strip in which one character printed the word banana in the sand and another character then slipped on the word and fell.

I think it is worth following up such bits of folk wisdom. People do indeed behave toward words and other symbols just as they behave toward their referents. Much of what we do is determined by things and events that we cannot possibly have experienced directly. For example, we can know historical events only through words; we can know most people only by what has been written about them or by their photographs and other representations, like statues; words and pictures make it possible for us to know about events that take place too far away for us to observe them directly; we deal effectively with quantities that are represented only by numbers on paper; we follow travel routes shown as lines on maps; and so on. The opportunity to apply a scientific analysis to such phenomena is, to me, more reason to develop and continue an interest in equivalence relations than is any possible personal theoretical triumph.

Each to his or her own, however. Those with a stronger theoretical bent will also find much of interest in the data on equivalence relations. (For different theoretical approaches, see, e.g., Hayes, Barnes-Holmes, & Roche, 2001; Horne & Lowe, 1996; Sidman, 1994, 2000.) But take care. One must judge theories on grounds of parsimony (precision, neatness, and simplicity), coherence, and consistency in their explanations of existing data, and productivity in their predictions of phenomena that have not yet been observed (see Sidman, 1997, particularly pp. 138–143). Finally, a theory must be capable of disproof. The methodology of theory construction is just as demanding as is the methodology of experimentation. Furthermore, do not go into theory construction under the illusion that you can escape from the technological constraints of rigorous experimentation. Failure to attend to the subtleties of experimental methodology will make one unable to evaluate rigorously the data that must inevitably be produced to test any theory.

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.

Carrigan, P. F., Jr., & 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.

Dube, W. V., Balsamo, L. M., Fowler, T. R., Dickson, C. A., Lombard, K. M., & Tomanari, G. Y. (2006). Observing behavior topography in delayed matching to multiple samples. *The Psychological Record, 56,* 233–244.

Dugdale, N., & Lowe, C. F. (2000). Testing for symmetry in the conditional discriminations of language-trained chimpanzees. *Journal of the Experimental Analysis of Behavior, 73,* 5–22.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York: Kluwer Academic/Plenum.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 185–241.

Iversen, I. H., Sidman, M., & Carrigan, P. (1986). Stimulus definition in conditional discriminations. *Journal of the Experimental Analysis of Behavior, 45,* 297–304.

Johnson, C., & Sidman, M. (1993). Conditional discrimination and equivalence relations: Control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 59,* 333–347.

Kastak, C.R., Schusterman, R., & Kastak, D. (2001). Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior, 76,* 131–158.

McDonagh, E. C., McIlvane, W. J., & Stoddard, L. T. (1984). Teaching coin equivalences via matching to sample. *Applied Research in Mental Retardation, 5,* 177–197.

McIlvane, W. J., & Dube, W. V. (2003). Stimulus control topography coherence theory: Foundations and extensions. *The Behavior Analyst, 26,* 195–213.

Ray, B. A., & Sidman, M. (1970). Reinforcement schedules and stimulus control. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules* (pp. 187–214). New York: Appleton-Century-Crofts.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 14,* 5–13.

Sidman, M. (1980). A note on the measurement of conditional discrimination. *Journal of the Experimental Analysis of Behavior, 33,* 285–289.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative.

Sidman, M. (1997). Equivalence: A theoretical or a descriptive model? *Mexican Journal of Behavior Analysis, 23,* 125–145.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74,* 127–146.

Sidman, M. (2008). Symmetry and equivalence relations in behavior. *Cognitive Studies, 15,* 322–332.

Sidman, M., & Cresson, O., Jr. (1973). Reading and transfer of crossmodal stimulus equivalences in severe retardation. *American Journal of Mental Deficiency, 77,* 515–523.

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., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52,* 261–274.

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review, 57,* 193–216.

Stoddard, L. T., Bradley, D. P., & McIlvane, W. J. (1987). Stimulus control of emergent performances: Teaching money skills. In J. A. Mulick & R. F. Antonak (Eds.), *Transitions in mental retardation: Vol. 2. Issues in therapeutic intervention* (pp. 113–149). Norwood, NJ: Ablex.

Vaughan, W., Jr. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 36–42.