

## NONAUTOMATED PROCEDURES IN DERIVED STIMULUS RELATIONS RESEARCH: A METHODOLOGICAL NOTE

SIMON DYMOND  
*APU, Cambridge*

RUTH ANNE REHFELDT  
*Southern Illinois University*

JACQUELINE SCHENK  
*Erasmus University, Rotterdam, The Netherlands*

Nonautomated or tabletop procedures are widely used in derived stimulus relations research. These procedures offer several advantages to the researcher, not least of which is the interactive format of the task. However, this feature is often criticized because of the possibility of experimenter cuing and imprecise experimental control over task presentations. These limitations, combined with the considerable procedural differences that exist between studies, suggest that a review of nonautomated procedures in derived stimulus relations research is warranted. The present paper will consider some of the methodological features of nonautomated procedures including experimental setting and sessions, task format, experimenter training, response definition, reinforcer delivery, intertrial intervals, and interobserver reliability. Basic methodological safeguards will be proposed to ensure that experimental control is rigorously maintained in future research with nonautomated procedures.

Ever since Sidman's (1971) pioneering study, research on derived stimulus relations has grown to become the most studied research area in the experimental analysis of human behavior (Dymond & Critchfield, 2001). The basic outcome of a study on derived stimulus relations is as follows. For instance, on hearing the spoken word "car," typically developing young children may be taught to point to an actual toy car. This training episode may be described as 'hear name, see and point to object.' Next, when presented with a picture of a toy car, a child may be instructed to point to the actual toy car ('see picture, point to object'). Following a training history like this, novel or emergent relations among the stimuli can be tested. That is, a child may produce the spoken word "car" and point to a picture when shown the actual toy car (called mutual

We are grateful to Paul Smeets and Bill Dube for providing helpful comments on an earlier version. Address correspondence to Simon Dymond, Department of Psychology, APU, East Road, Cambridge, CB1 1PT, UK. (E-mail: s.dymond@apu.ac.uk).

entailment), without further training. Similarly, a child may point to a picture on hearing the spoken word, and vice versa (called combinatorial entailment), all without additional training. Following such derived performances, the stimuli are said to participate in an equivalence class (Sidman, 1994) or relational frame of coordination (Hayes, Barnes-Holmes, & Roche, 2001).

Perhaps what is most interesting about derived stimulus relations such as equivalence is that these outcomes are not readily predicted from the traditional behavioral concept of conditional discrimination; neither the spoken word "car" nor the picture of a car has a direct history of differential reinforcement with regard to the other, and therefore neither stimulus should occasion the other. The processes responsible for these outcomes cannot be based upon stimulus generalization because the stimuli have no formal properties in common being from different modalities. Clearly, emergent performances such as these are critical components of language development and have considerable instructional significance (see Carr & Felce, 2000; O'Donnell & Saunders, 2003; Stromer, Mackay, & Remington, 1996).

In the literature on derived stimulus relations, a wide variety of experimental procedures have been employed with a host of different subject populations. These procedures can be broadly categorized as either automated or nonautomated (Saunders & Williams, 1998). Automated procedures typically involve computer-controlled stimulus presentations and recording of responses, while nonautomated procedures rely upon an experimenter to arrange trials and to deliver consequences. A brief inspection of the literature shows that of the 20 studies on derived stimulus relations with human subjects published in *The Psychological Record* in 2000, 9 (45%) used nonautomated procedures, while in the *Journal of the Experimental Analysis of Behavior*, 4 studies were published, 2 (50%) of which used nonautomated procedures. The majority of these studies have been conducted with typically developing children or individuals with developmental disabilities. Thus, it is apparent that tabletop procedures are extremely common in experimental investigations, and are certainly no less common than automated.

When a researcher is working with children or individuals with developmental disabilities, nonautomated procedures offer several advantages. These include the interactive nature of the tasks, the use of social reinforcers, and sessions of varying duration. Nonautomated procedures are also flexible and can be closely linked to regular ongoing instruction that might be occurring in an applied setting. Indeed, according to Pilgrim (1998):

Some research questions or environment-behavior interactions make automation impractical or even impossible, and human experimenters are required to either measure behavior, implement procedures, or both. Some response classes, particularly those involving more naturalistic behaviors, may be difficult to measure in the absence of a human observer.



However, each advantage brings with it additional difficulties that may threaten the internal validity of an experiment including the possibility of experimenter cuing effects and the need to continually assess an experiment's "procedural integrity" (Saunders & Williams, 1998, p. 195). It has been argued that these difficulties often mean that "researchers are more likely to question novel outcomes that are obtained with [nonautomated] procedures than those obtained under more automated procedures" (Saunders & Williams, 1998, p. 195). Given the widespread use of nonautomated procedures in derived stimulus relations research and the dispute surrounding their use, a review of the methodological features and possible drawbacks of such procedures seems warranted. In the present paper, we will provide suggestions, in addition to highlighting drawbacks, for how researchers using tabletop procedures can make the most of their protocols.

### Methodological Features of Nonautomated Procedures

In the interests of clarity, the following definition of nonautomated procedures will be adopted. *A nonautomated procedure is one in which all events are delivered, arranged, recorded, and consequence by an experimenter who is physically present in the experimental environment with the subject at all times.* The present paper will also consider research conducted with partially automated procedures in which stimulus presentations are controlled by computer and consequences are delivered by an experimenter. Nonautomated procedures, also referred to as tabletop procedures, typically involve the subject and experimenter sitting across a table from one another. To initiate a trial, the experimenter presents the stimuli in front of the subject and usually gives verbal instructions about the task. Following an appropriate response, the experimenter delivers verbal feedback and may instruct the subject to take a predefined reinforcer, records the response, and arranges the next trial. This methodological definition is not restricted to experiments on derived stimulus relations however; the assessment and treatment of behavior in applied settings has employed very similar arrangements for decades.

The present paper will now address methodological issues surrounding nonautomated procedures, including the experimental setting and task format, experimenter and observer training, response definition, reinforcer delivery, intertrial intervals, and assessing reliability. Then we will propose minimum methodological controls for future research and provide an illustrative research example of how to address any potential drawbacks.

#### *Experimental Setting and Sessions*

The researcher's first task in using a nonautomated procedure is to determine the experimental setting. An experimental setting that is distracting, cluttered, noisy, or has frequent interruptions can adversely affect a subject's performance on test trials as well as hindering

acquisition. The minimal requirements for a laboratory or university setting should be a relatively small, quiet room containing a table and at least two chairs. The subject and experimenter should sit either facing each other on opposite sides of the table or adjacent to each other facing the stimulus display. The furniture should be comfortable, in decent condition, and adjusted to the approximate height and size of the subject. It may be difficult to ensure the availability of such a research space in a school, clinic, or other service setting, but the researcher should seek to ensure regular access to something that approximates this goal, such as an unused classroom or playroom which is free from immediate visual and auditory distraction (see Pilgrim, 1998). The researcher should do what he or she can to prevent interruptions from others and to make the setting as comfortable and natural as possible, particularly when subjects are children or individuals with developmental disabilities. The more natural and comfortable the setting is for such subjects, the more likely they will be to perform optimally. Decorating bulletin boards in the room with children's cartoon figures, for example, may put younger subjects at ease.

Several experimental sessions per subject are typically conducted in derived stimulus relations research. This is due to the fact that several sessions are often necessary for subjects to acquire the baseline conditional discriminations, and because some subjects may fatigue after 20-30 min. For this reason, some researchers might schedule at least three sessions a week per subject, with each session lasting no longer than 20-30 min. Subjects should also be given breaks when necessary. Optimal performance is unlikely when a subject is physically uncomfortable, tired, or irritated. Finally, the experimenter should treat subjects with courtesy and convey gratitude for the individual's participation (Pilgrim, 1998).

Pilgrim, Jackson, and Galizio (2000) reported the use of a nonautomated procedure to assess the acquisition of conditional discriminations in 3- to 6-year-old typically developing children. Experimental sessions were conducted at the children's preschool or after-school program. Four to five sessions were conducted per week, each of which lasted approximately 15-20 min (Pilgrim et al., 2000).

### *Task Format*

The next important consideration concerns the format of the task itself. There are a number of seemingly unimportant variables in nonautomated procedures that, if overlooked, can potentially exert control over a subject's behavior. These include the positioning of the comparison stimuli, the distance of the comparison stimuli from the subject, and the possibility of the subject viewing the experimenter's datasheet. Fortunately such potential confounding variables are easily handled. A researcher can take certain precautions to ensure that a subject's performance is not differentially affected by the format of the task.

Research on derived stimulus relations is typically conducted using a match-to-sample (MTS) format. With this format, the presentation of a



sample stimulus marks the onset of each trial. If the sample stimulus is of the visual modality, a notecard displaying the sample stimulus may be handed to the subject or placed on the table directly in front of the subject. If the sample stimulus is of the auditory modality, the production of the experimenter's spoken word marks the onset of each trial. Following the presentation of the sample stimulus, two or more comparison stimuli are placed in a horizontal row, evenly spaced, on the table directly below the sample stimulus. To ensure that the stimuli are equidistant from each other as well as from the subject, some researchers present the sample and comparison stimuli together on one laminated sheet of paper, with each new trial represented by a new sheet. This way, the distance between the stimuli can be held consistent across trials. This procedure is also desirable because the left, right, and center positions of three comparison stimuli can be randomized across trials, thus preventing the position of a particular comparison stimulus from acquiring control over the subject's behavior. The consistency of the distance between the stimuli can also be ensured by the use of a placement board, in which the placement of each of the stimuli is clearly indicated on a piece of poster board or cardboard.

Subjects may be instructed to "match" or "put with the same" in studies using nonautomated procedures. For typically developing subjects it is usually not necessary to repeat such instructions beyond the beginning of a session. Sometimes an observing response to the sample stimulus is required, in which the subject is to point to, touch, or name the sample stimulus, in order to initiate the presentation of the comparison stimuli. This requirement ensures that the subject has made perceptual contact with the sample stimulus. The response requirement may also vary. In some studies subjects have been required to place the sample stimulus on top of the matching comparison, pick up and select a matching comparison stimulus, or point to their selection. The experimenter then delivers verbal praise or verbal praise accompanied by a tangible item if the subject is a child or an individual with a developmental disability. Incorrect trials are often repeated until the subject responds correctly.

Occasionally, with task arrangements of this kind, it may be difficult to prevent subjects from either seeing the experimenter's data sheet or from obtaining indirect feedback about their performance by discriminating the different scoring actions made to "correct" and "incorrect" responses. In such cases, many researchers choose to either employ a second observer to record all responses, to videotape the sessions and record responses later, or to adopt a clear and unambiguous data coding strategy. We will return to this point in another section. It is also important to ensure that the seating arrangements of the subject and experimenter ensure unobtrusive data recording and task presentation. Stimuli should be presented from a convenient, hidden position so that subjects cannot foresee or predict the next trial. In situations where this is either not possible or desirable, tasks can be presented on bound index cards, for

example, with which subjects can initiate and present each trial themselves (e.g., de Rose, McIlvane, Dube, Galpin, & Stoddard, 1988). Also, tasks should require as little response effort as possible and be tailored to the developmental and behavioral capacity of the subject. For instance, one of the reasons that the MTS format is considered appropriate for use in derived stimulus relations research with young children is because of its similarity to educational exercises in which relations are taught between objects and their corresponding written or spoken names (Hayes et al., 2001). Some authors have sought to approximate this developmental history in research with young children by incorporating within their MTS procedures features of real-world matching exercises including thematic or taxonomic stimuli (e.g., Osborne & Calhoun, 1998), training of object-name relations (e.g., Brady & McLean, 2000; Lipkens, Hayes, & Hayes, 1993), and the response of physically placing a sample stimulus upon a comparison (e.g., Smeets, Barnes-Holmes, Akpinar, & Barnes-Holmes, 2003; Smeets, Schenk, & Barnes, 1995). DeGrandpre, Bickel, and Higgins (1992) used a nonautomated procedure to assess the establishment of equivalence relations between interoceptive drug stimuli and exteroceptive visual stimuli. In this study, training and testing of stimulus relations were conducted with the subject and experimenter sitting and facing one another with a card table between them. On visual-visual trials, the subjects were instructed to "match the visual stimulus with the visual stimulus" (DeGrandpre et al., 1992, p. 13). Subjects were to indicate their response by answering "left," "right," or "center" while simultaneously pointing to the stimulus of their choice. All of the stimuli were presented simultaneously on a placement card directly on the table in front of the subjects.

In a partially automated procedure, Goyos (2000) employed a computer to arrange and record stimulus presentations and to deliver auditory feedback for correct and incorrect responses. The subject responded by pointing to a stimulus displayed on the computer screen. A video camera and monitor allowed the experimenter to observe the child's behavior from an unobstrusive position such that when the child pointed to a stimulus, the experimenter pressed the appropriate key to record the response and to activate the auditory feedback.

### *Experimenter Training*

A critical yet often overlooked feature of nonautomated procedures involves experimenter and observer training. According to Saunders and Williams (1998, p. 195), "the most important concern is that the experimenter might inadvertently prompt or provide feedback to the subject. Even subjects with extreme developmental limitations bring to the laboratory a long history of following nonverbal prompts. Moreover, it is surprisingly difficult for many experimenters to suppress inadvertent cues, especially premature motions toward delivering consequences" (p. 195). Such cues, be they subtle gestures or facial expressions, are



particularly problematic on test trials, for it is possible that any emergent performances observed were established by prompts. On training trials such prompts or cues prevent a subject's responding from coming under control of the reinforcement contingencies.

Clearly, the interactive nature of nonautomated procedures brings with it the possibility of inadvertent cues from the experimenter influencing the behavior of the subject. In order to limit possible experimenter cuing effects, it is important to closely monitor the behavior of both the experimenter(s) as well as any other observers. Recommended guidelines for the training of observers can be found in several textbooks on behavioral research methods (e.g., Johnston & Pennypacker, 1993; Pilgrim, 1998; Poling, Methot, & LeSage, 1995) and can be applied to experimenter training practices prior to the start of a nonautomated study. For instance, Pilgrim (1998) suggested that experimenter training "might include modelling or role-play with feedback; videotapes can be used for practice with measurement or with identifying occasions for reinforcement. Training should continue until the trainee consistently meets predetermined accuracy criteria under conditions as similar to those of the experiment as possible" (p. 32). Role-playing with feedback, where a senior researcher assumes the role of a subject, is a particularly useful strategy for training experimenters because it mimics experimental conditions and allows for the immediate identification and resolution of any problems.

In general, there should be three objectives to any experimenter training procedure. First, details of the experimental contingencies need to be comprehensively explained and demonstrated. Second, procedures for maintaining on-task behavior need to be outlined. In research with young children, for instance, potential problems such as attempting to look at the data sheet, impulsive or inadequate responding (e.g., pointing to the stimuli with tongue or elbow, at all comparisons simultaneously, or without looking), and noncompliance (e.g., not responding or claiming not to know the answer) can be identified through role-playing. Self-stimulatory behavior (e.g., waving or flapping the stimuli) may be a source of concern among subjects with severe developmental disorders. Researchers should attempt to ensure that inappropriate behaviors are reduced through the scheduling of brief, daily experimental sessions and the provision of supplemental reinforcers contingent upon attendance at all scheduled sessions and appropriate on-task behavior. Third, the prevention of facial cues (e.g., eye-darting) by the experimenter should be targeted. Although brief facial cues from an experimenter may cue the predicted response, instructing experimenters to stare at a fixed location on the stimulus display or on his or her data sheet readily prevents any confounding influence over the subjects' performance.

A study by Rehfeldt, Latimore, and Stromer (2003) established derived stimulus relations among subjects with developmental disabilities. Subjects were also probed for the emergence of derived stimulus relations with stimuli with which a peer had modeled conditional

discriminations. Experimenters were staff who were employed in an after-school discrete trial teaching program for children with autism. As such, the skills necessary to serve as experimenters were similar to the skills required for their jobs. Skills were established using common behavior analytic approaches to staff training (e.g., Reid & Green, 1990). Training consisted of staff first being provided with a verbal description of the desired performance, the desired performance was then modeled by an advanced therapist, staff were then given constructive feedback using a checklist similar to that shown in Table 1, and so on, until the desired discrete trial teaching skills were mastered. Role rehearsal training was also used.

Table 1

An Example of a Staff-Training Checklist Used by Rehfeldt et al. (2003)

	Yes	No	N/A
<b>PRESESSION</b>			
1. Consults preference assessment outcomes			
2. Selects developmentally appropriate skills from Maurice book			
3. Has therapy materials and data sheets ready			
4. Creates data sheet for therapy session			
5. Has reinforcers ready			
6. Cleans environment of distractions			
Total # Correct: _____			
<b>SESSION</b>			
1. Has client's attention			
2. Presents task			
3. Uses minimal instructions			
4. Uses prompt to guarantee the desired behavior will be emitted			
5. Delivers reinforcers for correct behavior			
6. Delivers reinforcers immediately			
7. Uses enthusiasm when delivering reinforcers			
8. Uses appropriate amount of reinforcement			
9. Scores data promptly			
10. Begins with easier tasks			
11. Intersperses difficult with easy tasks			
12. Maintains rapid presentation of task			
13. Gives child necessary breaks			
14. Ends session with easy fun task			
Total # Correct: _____			

### *Response Definition*

The response topographies in conditional discrimination research using automated or nonautomated procedures vary widely. In MTS studies using automated procedures, subjects are usually required to select a stimulus by touching (using one of their fingers) stimuli that are presented on a touch screen, by moving and clicking a mouse (e.g., Jordan, Pilgrim & Galizio, 2001) or pressing stimulus display-response keys (e.g., Stoddard, 1982; Stromer, MacKay, McVay, & Fowler, 1998). In studies using nonautomated procedures, subjects are usually required to point to and touch (using one of their fingers) stimuli that are presented



on a laminated card showing the sample and/or comparison stimuli (e.g., Dube, McIlvane, Mackay, & Stoddard, 1987; Schenk, 1993) or stimuli that are presented on a wooden display with comparison stimuli in predetermined positions (Dugdale & Johnson, 2002). In the study by Dugdale and Johnson (2002), for example, comparison stimulus cards were placed in predetermined positions on wooden leaves of a 'book,' which was placed on the table facing the 2-year-old subject, who was then asked to touch or point to one of the comparison stimuli.

According to Saunders and Williams (1998), one of the main drawbacks of nonautomated procedures is that "immediate decisions as to whether responses meet the experimental contingencies may be difficult. For example, a subject may barely touch one stimulus and then move quickly to another" (p. 195). Although invalid responding may occur both in studies using automated as well as nonautomated procedures, it is implied that concerns regarding whether the observed responses meet the experimental contingencies mainly apply to nonautomated procedures. For instance, unless care is taken to ensure an automated procedure is programmed such that trials are only terminated when a subject has responded to one of the comparisons, invalid responding may still occur (e.g., by pressing all response keys simultaneously). Also, as Saunders and Williams agree, any such difficulties can easily be controlled by careful response definition or "by requiring a more definite response such as sorting the cards or handing the cards to the experimenter" (1998, p. 195). Stimulus sorting tests involve subjects receiving all the stimuli and being instructed to place the objects into groups, in a way they think is most appropriate (e.g., Pilgrim & Galizio, 1996). For example, in a study by Smeets, Dymond, and Barnes-Holmes (2000), subjects were presented with a sheet containing the instruction to categorize stimuli into two groups and that "if you find £ and # belong to one group, put an X next to £ and # in Group 1, and an X next to the other stimuli in Group 2" (p. 346).

Other forms of responding can be required as alternatives to pointing and touching stimuli. For instance, instead of printing stimuli directly on a card, researchers could apply a nonpermanent adhesive material such as Velcro®, to the front or back of comparison stimulus cards and require subjects to place the sample stimulus card onto the comparison stimulus card (see also, Bush, 1993). Using Velcro®, it is possible to train stimulus relations by ensuring that only one sample card (e.g., A1) would "stick to" the correct comparison (B1) and not to the incorrect comparison (B2), giving the subject immediate feedback on his or her choices. For test trials, one would consider 'baiting' both comparison cards identically and preventing any guidance as to which choice might be accurate. Also, the use of magnetized stimulus materials could be considered. A magnetized display board could be attached to a wall or screen, and the experimenter could place one sample stimulus magnet onto the display board and require the subject to place one of the comparison magnets next to the sample magnet. In this way, the required form of responding would be highly comparable to pointing to and clicking with a mouse in automated procedures.

Other, more definite forms of responding are illustrated by examples of nonautomated studies in which responses are defined as clapping and waving hands (Smeets, Barnes, & Roche, 1997), selecting one of two objects (Dube et al., 1987; Smeets & Barnes-Holmes, 2003), and physically picking up and placing the sample stimulus onto one of the comparison stimuli (Smeets et al., 1995). In these studies, making quick decisions as to whether the subject has made the predefined response or not is facilitated because of its more definite nature.

Not unlike automated experiments, researchers using nonautomated procedures must take the issues of response definition and how to assess whether a response meets the experimental contingencies very seriously. This is particularly the case when subjects are required to touch or point to stimuli. Clear definitions are used of both valid (observing responses, pointing to and touching one stimulus using one finger) and invalid responding (e.g., pointing at the stimulus with eyes closed, pointing at both stimuli, hovering above one comparison stimulus for a number of seconds with or without touching it, pointing at stimuli with nose, elbow, or fist, etc.). Next to having clear definitions, researchers also have clear guidelines for how to respond to valid responding (i.e., verbal approval and delivery of a token in training trials where valid responding was accurate, and verbal disapproval and no delivery of a token in cases where valid responding was inaccurate) and to invalid responding (i.e., a corrective remark ["No, that's not the correct way to pick a shape"], followed by a single demonstration by the experimenter of how to respond in a valid manner with no reference to the designated S+). On test trials, invalid responses, should they occur, can either be corrected or ignored by the experimenter. Researchers can be encouraged by high levels of agreement between raters, which is indicative of clear and concise response categories provided responses are distributed across all of the categories.

Ways of overcoming problems of response definition with MTS procedures may entail incorporation of some or all of the following guidelines. First, researchers should clearly define the critical area surrounding the comparison stimuli and record responses made only within this area. Response calibration can easily be achieved by, for example, displaying stimuli within text boxes and by only recording pointing responses within these boxes (e.g. Schenk, 1993, 1995; Smeets et al., 1995). Second, researchers should define the response ratio requirements such as permitting only one touch of the stimulus display before feedback is delivered. Researchers should also adopt clear guidelines for how to attempt to correct invalid responding such as touching a stimulus more than once, touching more than one stimulus by, for example, having standardized feedback (e.g. "No! You should point like this" [demonstrate pointing]). Third, all responses both on and off task including verbal behavior should be recorded since it is common in research with young children for off-task behaviors and response latency to decrease over trials as experimental control over the subjects' behavior



is acquired. This information, including any verbalizations or comments on the task and task materials might provide valuable information regarding the nature and strength of any derived relations (Dymond & Rehfeldt, 2001). Fourth, the responses required might involve topographies such as clapping or waving hands which would clearly enhance the observation and recording of behavior (Barnes, Browne, Smeets, & Roche, 1995). Finally, it should be ensured either formally or informally before the start of the experiment that subjects possess the necessary fine-motor and/or receptive language skills that are required for emitting the responses, that they can understand instructions provided by the experimenter, and that this level of ability is comparable across subjects. This might even apply more to studies using automated procedures, particularly when subjects are required to choose from a selection of keys or to use a computer mouse for stimulus selection

In a study by Smeets et al. (1995), comparison stimuli were presented on a standing display board and subjects were required to physically place the sample stimulus card, which was put in front of the display board by the experimenter, on one of the comparison cards (see Figure 1). Once subjects had responded by placing the sample card onto one of the comparison stimuli, the experimenter placed another sample card in front of the display, which subjects then had to put with the other comparison stimulus card on the display. Only thereafter would the experimenter remove all stimulus cards and begin with the next trial. With this form of responding, the decision as to whether a subject's responses have met experimental contingencies does not leave much room for any subjective interpretation by the experimenter or observer. Determining whether a subject has *pointed* to a comparison stimulus is perhaps more open to subjective interpretation by the experimenter than deciding whether a subject has *physically placed* a stimulus card onto a display board. This might be even more important when the subject has to select one stimulus from multiple comparisons.

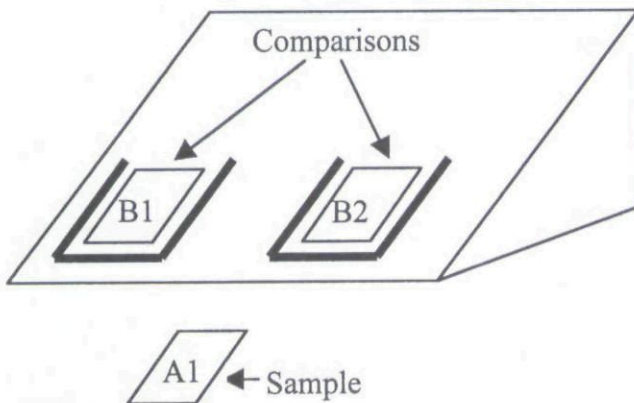


Figure 1. Display used for MTS tasks in Smeets, Schenk, and Barnes (1995). The stimulus cards on the display represent comparison stimuli, the stimulus card in front of the display represents the sample stimulus. The display is shown from the child's perspective.

### *Reinforcer Delivery*

The interactive nature of nonautomated procedures is one of their main advantages in derived stimulus relations research with normally developing children, children with developmental disabilities, and adolescents because it resembles the everyday learning situations of subjects, involving a context with one or two adults present (e.g., school, home environment) with comparable consequences for behavior. In studies using nonautomated procedures, consequences such as verbal praise, edibles such as biscuits, cheese biscuits, M&Ms, and tokens are typically used for accurate responding during training and are either hand-delivered by the experimenter or selected by the subject (e.g., Dube et al., 1987; Pilgrim et al., 2000; Schenk, 1993, 1994; Smeets et al., 1995). However, "because reinforcers are hand-delivered, there is variation in the amount of time between response and reinforcer delivery" (Saunders & Williams, 1998, p. 195). While the effects of delayed reinforcement on operant acquisition are well documented (e.g., Lattal & Gleeson, 1990), comparable analyses for derived relational responding have yet to be undertaken. It remains an empirical issue, therefore, whether any variation in the amount of time between response and reinforcer delivery facilitates or impedes performance in derived stimulus relations research.

Compared with automated procedures, where there is immediate response-reinforcer contiguity, what is observed in nonautomated reinforcement procedures is often as follows. Firstly, at the beginning of a session the researcher may explain the "rules of the game" to a child and give her a choice from an array of preselected reinforcers such as stickers, books, and small toys. This toy or prize is then placed in view of the child, usually at the corner of the table, and access is only allowed following completion of a fixed-ratio number of correct responses with secondary reinforcers such as colored beads or marks on a subject registration form. The fixed-ratio requirement can vary between 30, 50, and even 100 correct responses, with larger ratios being employed with older children. It is then explained to a child that he or she can win the preselected toy only after all of the beads are transferred from one container to another. The researcher delivers verbal praise ("Good!," "Well done!," "Excellent!") and instructs the child to take a bead immediately following each correct response. Incorrect responses typically result in negative feedback (i.e., "No, that is wrong. No bead."). Also no beads are allowed for invalid responding (e.g., Boelens, 1990; Schenk, 1993, 1994).

These reinforcement arrangements confer several advantages. First, a range of preferred, preselected reinforcers are available, ensuring that the child's behavior is under reliable reinforcement control. Second, the schedule of reinforcer delivery can be readily altered to suit the individual experiment and avoid problems of satiation. Third, social reinforcers can be delivered noncontingently to maintain on-task behavior and active participation in the task (e.g., Pilgrim et al., 2000). Fourth, the child has



concurrent visual access (during training) to the accumulating reinforcers and "may work harder for tokens, exchangeable for prizes, when those prizes are visible throughout the experimental session" (Pilgrim, 1998, p. 34). Finally, the delivery of verbal praise is relatively immediate and makes use of a range of conditioned reinforcers (i.e., "that's right!," "good job!," etc). The use of verbal praise may be of particular importance to children because of verbal praise playing such a prominent role in class management, parenting styles, and home treatments emphasising positive child-adult interaction (Weaver, Watson, Cashwell, Hinds, & Fascio, 2003).

In studies using nonautomated procedures, criteria for the exchange of tokens (e.g., beads) earned by the subject during training can vary from exchange contingent upon a certain number of tokens to exchange contingent upon participation during each scheduled session (Pilgrim et al., 2000). For example, in the study by Pilgrim et al. (2000), following the last session of the week, subjects were permitted to choose a small present (a toy) from a 'treasure chest,' contingent on participation during each scheduled session.

Schenk (1994) used a nonautomated procedure to examine the effects of outcome-specific consequences on formation of equivalence relations. On every training trial pointing to the designated correct comparison was followed by the experimenter saying "Good! Take a blue/red bead." The experimenter then removed all stimulus materials and waited until the subject had placed the bead into the accompanying tube of the same color (blue/red) before beginning the next trial. Filling the tube required 25 beads and children were then given their preselected picture or toy, the tube emptied, and the study resumed.

Lionello-DeNolf and McIlvane (2003) reported a recent laboratory description of an automated version of the Wisconsin General Test Apparatus (WGTA) for use with individuals with developmental disabilities. In this study, the opening or closing of food compartments was computer-controlled and the compartments were also equipped with sensors to indicate when a subject touched them, allowing the researchers to standardize reinforcer delivery times and to record response latency.

Boelens, van den Broek, and Calmeyn (2003) employed a partially automated procedure with normally developing young children (Experiments 2 and 3). On an auditory-visual MTS training trial, the experimenter first spoke the sample name and then visual comparison stimuli were displayed on either side of an area in which the computer mouse cursor was stationary. Subjects were required to move the cursor to one of the comparisons which, following correct responses, resulted in the screen being cleared and the presentation of a white rectangle, a colored star, and a brief (3 s) musical jingle. The experimenter then presented the child with a token and when 40 tokens were accumulated the child was given a preselected picture which was glued into an individualized display book, and was allowed to choose another picture.

In the partially automated procedure of Goyos (2000), correct responses resulted in auditory feedback from both the computer and the

experimenter and a token delivered into a tray. The subject could then retrieve the token and place it in one of two positions on a matrix that permitted concurrent visual access to the accumulated reinforcers.

### *Intertrial Intervals*

In nonautomated procedures, intertrial intervals (ITI) may vary from one trial to the next and hence, "it may be difficult to control or measure aspects of trial timing, such as the length of time between trials or response latency. In fact, the exact beginning of a trial may be somewhat ambiguous if care is not taken to prevent the subjects from viewing the stimuli before they are completely positioned" (Saunders & Williams, 1998, p. 195). That ITIs are likely to vary throughout a nonautomated study need not be considered a drawback for the following reasons.

First, a standardized ITI, although desirable, is not an absolute prerequisite in nonautomated studies. If necessary, trials can be presented in intervals of time to match a subject's pace and level of responding. That is, more trials can be given in relatively quick succession when a subject is responding at a consistent and correct high rate. Second, it is unlikely that response latency would be employed as a supplemental measure of derived stimulus relations in a nonautomated study because it is necessary to have fixed ITIs in order to accurately record response onset and offset (Dymond & Rehfeldt, 2001). Undertaking such an analysis of response speed is simply unfeasible with a nonautomated procedure. Third, it is possible to measure ITIs post hoc by analyzing real-time video recordings of experimental sessions where simple observer reliability checks can then be performed (e.g., Miltenberger, Rapp, & Long, 1999). One derived stimulus relations experiment with typically developing young children found that ITIs ranged between 5 s and 20 s—substantially longer than similar automated procedures (Dymond, 2000). Over the course of the study, however, ITIs stabilized (*mean*: 8 s) as the behavior of subject and experimenter showed a gradual adjustment to the procedures. Finally, the statement that the exact beginning of a trial may be difficult to determine because the subjects may view the stimuli before they are positioned is easily countered by ensuring that all tasks are presented in such a way as to prevent the subjects seeing the stimuli before they are displayed, or allowing them to present tasks themselves, on bound index cards, for instance. While a degree of uncertainty is likely to surround the precise beginning of a nonautomated trial, it is not sufficient to jeopardize experimental control or make interpretation of findings obtained with it problematic assuming that these minimal safeguards are followed.

Although it is uncommon for studies employing nonautomated procedures to report aspects of trial timing, studies in which ITIs are reported often omit the procedure for counting them. In such studies where it may be necessary to maintain consistent ITIs, a simple and efficient method involves the experimenter covertly counting the interval



between stimulus presentations (Leader, Barnes-Holmes, & Smeets, 2000). For example, Leader et al. (2000) presented pairs of arbitrary stimuli on observation cards for 1 s each to young, normally developing children. Stimulus duration and within-pair delays of 1 s were estimated by the experimenter covertly counting "21." This simple strategy is one way of standardizing ITIs in a nonautomated procedure. Other strategies include the use of an unobtrusive digital timer or a hand-held computer set to chime at predetermined intervals, similar to the continuous recording methods used in applied settings (see Kahng & Iwata, 1998). Regardless of the method employed, researchers should seek to ensure that a consistent intertrial interval is arranged and that the procedure on which it is based is reported in replicable detail.

### *Interobserver Reliability*

The fact that all responses in nonautomated procedures are recorded by human observers makes it necessary to assess the reliability of those observers (Saunders & Williams, 1998, p. 195). Interobserver agreement in a nonautomated study is typically expressed as a percentage agreement score by dividing the number of trials in which both observers scored the same response by the total number of observed trials.

Undertaking assessments of interobserver agreement involves having either an additional observer present for some sessions or video recording all sessions. In studies employing an additional observer, it is customary for the observer to sit at an unobtrusive distance from the subject and experimenter where he or she can clearly see the subject's responses, but not the experimenter's data sheet. Later, the observer and experimenter should calculate a percentage agreement score for each session or experimental phase observed. With real-time video recordings, observers should be blind as to the predicted responses of subjects on each trial and agreement scores should be calculated for as many sessions of each experimental phase as possible (pretraining, pretesting, training, testing, follow-up, etc). An additional safeguard that can be employed with either of these approaches involves using a blind tester during critical test phases. Blind testers present and record all responses as an experimenter would but are typically only used during testing phases of a study. This controls for the possibility of experimenter cuing by having another 'experimenter,' blind as to the purpose of the study, present the tasks. Several nonautomated studies with multiple experiments have adopted and extended this practice by replicating findings with different experimenters (e.g., Smeets et al., 2003; Smeets & Barnes-Holmes, 2003), hence ruling out the possibility of experimenter cuing.

A brief inspection of the derived stimulus relations literature employing nonautomated procedures highlights both a wide range of minimum number of observed trials and range of acceptable interobserver agreement scores (some nonautomated studies have omitted reporting interobserver agreement scores at all (e.g., Carr, Wilkinson, Blackman, & McIlvane, 2000; Pilgrim et al., 2000)). For

instance, in their research with young children, Leader et al. (2000) assessed the reliability of 30% of all trials (training and testing) and reported an agreement score of 100%. In their study with adults with severe developmental disabilities and typically developing children, Brady and McLean (2000) undertook assessments of 13 videotaped sessions, including at least one session from each experimental phase, and reported agreement scores ranging from 87% to 100% with a mean of 97%. Overall, it is a good practical rule to calculate percentage agreement scores for all phases of a study and for at least 25% of total observations (Poling et al., 1995, p. 72). The range of acceptable interobserver agreement scores in applied behavioral research should be between 80% and 100% (Kazdin, 2001), whereas in derived stimulus relations research the suggested range should be between 90% and 100%. Scores lower than this threshold should be excluded from further analysis, the research suspended, and the practices used to train observers revised.

The reliability of observations is of paramount importance in all fields of science, not least those involving extended interaction between a subject and experimenter. Whether it be through the use of a second observer or real-time video-recordings of all sessions, calculating and reporting interobserver agreement is essential. Interestingly, although interobserver agreement is rarely reported in basic research when humans are used to assess the behavior of nonhumans (Poling, 1985), it is a useful shorthand to undertake assessments of reliability whenever human observers are employed.

### Other Nonautomated Procedures

The present paper has focused on nonautomated procedures which involve tabletop presentation of tasks and little specialist equipment. Other nonautomated procedures have been used in derived stimulus research, however, which it is argued offer a low-technology solution to many of the methodological issues described above (Saunders & Williams, 1998). Perhaps the most common alternative procedure is the Wisconsin General Test Apparatus (Lionello-DeNolf & McIlvane, 2003; Overman, Bachevalier, Turner, & Peuster, 1992; Pilgrim et al., 2000). The WGTA consists of a wooden box bisected with a guillotine door. The subject and experimenter sit on either side of the door which is raised during trials preventing visual contact between the two but allowing the experimenter to observe the subject's response. Trials are arranged on a tray with the door in the down position. The stimuli used with the WGTA are usually three-dimensional objects placed on the tray and slid towards the subject through the raised guillotine door. Typically, reinforcers such as edibles and money are arranged in "bait wells" under the stimuli and subjects respond by displacing the stimuli.

Saunders and Williams (1998) describe several advantages of the WGTA. These include the prevention of facial cues during the preparation



and presentation of trials, the delivery of reinforcement "at a uniform time with respect to the response" and "the ease with which three-dimensional stimuli can be used" (p. 195). For these reasons, the WGTA may be a desirable choice of apparatus for researchers conducting nonautomated research. However, many advantages afforded by the WGTA may also be achieved with other nonautomated procedures. We will now briefly consider each of these three advantages in turn.

First, the likelihood of the experimenter cuing with a tabletop procedure is reduced in the same manner as seen with the WGTA by arranging a free-standing divide between the subject and experimenter. For example, a curtained wooden frame can be assembled between the subject and experimenter preventing visual access to the stimulus array during trial preparation and, if additional observers are present, the curtain can be drawn during trial presentations to completely remove the possibility of experimenter cuing. Such a feature can be easily incorporated into other nonautomated and partially automated procedures.

Second, while the WGTA facilitates uniformity in reinforcer deliveries, uniformity across trials may not always be accomplished. In the majority of research conducted using the WGTA, edibles or money are used as reinforcers and are placed in the bait wells beneath the correct comparison. Each trial typically involves presenting the stimulus tray with a three-dimensional sample stimulus covering the center food well. The subject is instructed to displace the sample and the tray is then removed and presented again with the comparisons covering the wells on either side of the sample. Selection of the correct comparison reveals the reinforcer and is accompanied by verbal praise from the experimenter. Additionally, a research assistant often delivers noncontingent praise to maintain on-task behavior. Thus, the same challenges inherent in other nonautomated procedures regarding the timing of reinforcer deliveries may also be a source of concern using the WGTA. While the likelihood of a delay occurring between a response and a reinforcer must be minimized at all times regardless of the procedure employed, it remains an empirical issue whether or not delayed reinforcement affects the acquisition of conditional discriminations using nonautomated procedures *per se*. As outlined in an earlier section of this article, nonautomated procedures should arrange reinforcer deliveries to be immediately contingent on accurate responding and employ a range of generalized conditioned reinforcers to maintain on-task behavior.

Finally, although the WGTA lends itself well to investigations involving three-dimensional stimuli, such stimuli can be incorporated into other nonautomated procedures as well. For instance, in a nonautomated study conducted by Tierney, De Lary, and Bracken (1995) it was shown that adults could derive equivalence relations between visual and haptically perceived stimuli. The haptic stimuli consisted of three different wooden shapes (rectangles and a triangle) of the same weight and all but 1 of the subjects subsequently demonstrated equivalence relations. Similarly, visual and tactile access to three-dimensional stimuli is readily employed in nonautomated research with young children (e.g., Bush, 1993).

Thus, the advantages afforded by the WGTA may also be realized in other types of nonautomated procedures. One potential limitation to the WGTA that warrants mention is the replicability of results using other automated and nonautomated procedures. Future research will be necessary to resolve this concern (see Garrotti, DeSouza, DeRose, Molina, & Gil, 2000; Smeets et al., 2003).

### *Conclusion*

The wide range of derived stimulus relations research conducted using nonautomated procedures has led to uncertainty concerning the effectiveness of the various procedures used to facilitate derived relational responding and to the possibility that any resulting differences in outcomes may be caused by experimenter cuing, rather than the training and testing protocols employed (e.g., Leader et al., 2000; Pilgrim et al., 2000; Smeets & Striefel, 1994; Zygmunt, Lazar, Dube & McIlvane, 1992). Clearly this uncertainty and the questioning of results obtained with nonautomated procedures is not beneficial for the study of derived stimulus relations and for the development of new and interesting experimental procedures. However, adherence to a series of methodological controls, such as those outlined in the present paper, makes the possibility of experimenter cuing far less likely to influence nonautomated outcomes. Future nonautomated research should, therefore, seek to ensure that key methodological safeguards are adopted and vigorously enforced.

Adopting the methodological controls outlined in the present article could lead to increased behavioral research on those unique questions of behavior-environment interaction which make automation difficult, impossible, or unnecessary. For instance, novel procedures to empirically investigate derived stimulus relations of "more-than" and "less-than" with young children may involve tabletop presentation of familiar stimuli such as coins of different value. Providing that the appropriate safeguards are implemented, such procedures can help cast light on the development of this complex behavior (Hayes et al., 2001). The advantages of nonautomated procedures for studying derived relational responding in a naturalistic context with a variety of different subject populations and preparations must be fully utilized without sacrificing experimental control.

### References

- 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.
- BOELEN, H. (1990). Emergent simple discrimination in children. *Behavioural Processes*, 22, 13-21.
- BOELEN, H., VAN DEN BROEK, M., & CALMEYN, S. (2003). Is children's symmetric matching-to-sample the product of experience with spoken names? *The Psychological Record*, 53, 593-617.



- BRADY, N. C., & MCLEAN, L. K. S. (2000). Emergent symbolic relations in speakers and nonspeakers. *Research in Developmental Disabilities*, 21, 197-214.
- BUSH, K. M. (1993). Stimulus equivalence and cross-modal transfer. *The Psychological Record*, 43, 567-584.
- CARR, D., & FELCE, D. (2000). Application of stimulus equivalence to language intervention for individuals with severe linguistic disabilities. *Journal of Intellectual and Developmental Disability*, 25, 181-205.
- CARR, D., WILKINSON, K. M., BLACKMAN, D., & MCILVANE, W. J. (2000). Equivalence classes in individuals with minimal verbal repertoires. *Journal of the Experimental Analysis of Behavior*, 74, 101-115.
- DEGRANDPRE, R. J., BICKEL, W. K., & HIGGINS, S. T. (1992). Emergent equivalence relations between interoceptive (drug) and exteroceptive (visual) stimuli. *Journal of the Experimental Analysis of Behavior*, 58, 9-18.
- DE ROSE, J. C., MCILVANE, W. J., DUBE, W. V., GALPIN, V. C., & STODDARD, L. T. (1988). Emergent simple discrimination established by indirect relation to differential consequences. *Journal of the Experimental Analysis of Behavior*, 50, 1-20.
- DUBE, W. V., MCILVANE, W. J., MACKAY, H. A., & STODDARD, L. T. (1987). Stimulus class membership established via stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior*, 47, 159-175.
- DUGDALE, N., & JOHNSON, S. (2002). Unreinforced conditional selection by two-year olds in a six-comparison matching task. *The Psychological Record*, 52, 159-172.
- DYMOND, S. (2000, March). *Nonautomated procedures in derived stimulus relations research with children*. Paper presented at the annual conference of the Experimental Analysis of Behavior Group, London.
- DYMOND, S., & CRITCHFIELD, T. S. (2001). Neither dark age nor renaissance: Research and authorships trends in the experimental analysis of human behavior (1980-1999). *The Behavior Analyst*, 24, 241-253.
- DYMOND, S., & REHFELDT, R. A. (2001). Supplemental measures of derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin*, 19, 8-12.
- GAROTTI, M., DESOUZA, D. G., DEROSE, J. D., MOLINA, R. C., & GIL, M. S. A. (2000). Reorganization of equivalence classes after reversal of baseline relations. *The Psychological Record*, 50, 35-48.
- GOYOS, C. (2000). Equivalence class formation via common reinforcers among preschool children. *The Psychological Record*, 50, 629-654.
- HAYES, S. C., BARNES-HOLMES, D., & ROCHE, B. (2001). *Relational Frame Theory: A post-Skinnerian account of human language and cognition*. New York: Kluwer Academic Press.
- JOHNSTON, J. M., & PENNYPACKER, H. S. (1993). *Strategies and tactics of human behavioral research* (2nd ed.). Hillsdale, NJ: Erlbaum.
- JORDAN, C. R., PILGRIM, C., & GALIZIO, M. (2001). Conditional discrimination and stimulus equivalence in young children: Comparison of three baseline training procedures. *Experimental Analysis of Human Behavior Bulletin*, 19, 3-7.
- KAHNG, S., & IWATA, B. A. (1998). Computerized systems for collecting real-time observational data. *Journal of Applied Behavior Analysis*, 31, 253-261.
- KAZDIN, A. E. (2001). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.
- LATTAL, K. A., & GLEESON, S. (1990). Response acquisition with delayed reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes*, 16, 27-39.

- LEADER, G., BARNES-HOLMES, D., & SMEETS, P. M. (2000). Establishing equivalence relations using a respondent-type training procedure. III. *The Psychological Record*, 50, 63-78.
- LIPKENS, R., HAYES, S. C., & HAYES, L. J. (1993). Longitudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology*, 56, 201-239.
- LIONELLO-DENOLF, K. M., & MCILVANE, W. J. (2003). Rebirth of the Shriver Automated Teaching Laboratory. *Experimental Analysis of Human Behavior Bulletin*, 21, 12-17.
- MILTENBERGER, R. G., RAPP, J. T., & LONG, E. S. (1999). A low-tech method for conducting real-time recording. *Journal of Applied Behavior Analysis*, 32, 119-120.
- O'DONNELL, J., & SAUNDERS, K. J. (2003). Equivalence relations in individuals with language limitations and mental retardation. *Journal of the Experimental Analysis of Behavior*, 80, 131-157.
- OSBORNE, J. G., & CALHOUN, D. O. (1998). Themes, taxons, and trial types in children's matching to sample: Methodological considerations. *Journal of Experimental Child Psychology*, 68, 35-50.
- OVERMAN, W. H., BACHEVALIER, J., TURNER, M., & PEUSTER, A. (1992). Object recognition versus object discrimination. Comparison between human infants and infant monkeys. *Behavioral Neuroscience*, 106, 15-29.
- PILGRIM, C. (1998). The human subject. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 15-44). New York: Plenum.
- PILGRIM, C., & GALIZIO, M. (1996). Stimulus equivalence: A class of correlations or a correlation of classes? In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 173-195). North Holland: Elsevier Science.
- PILGRIM, C., JACKSON, J., & GALIZIO, M. (2000). Acquisition of arbitrary conditional discriminations by young normally developing children. *Journal of the Experimental Analysis of Behavior*, 73, 177-193.
- POLING, A. (1985). Reporting interobserver agreement: Another difference in applied and basic behavioral psychology. *Experimental Analysis of Human Behavior Bulletin*, 1, 5-6.
- POLING, A., METHOT, L. L., & LESAGE, M. G. (1995). *Fundamentals of behavior analytic research*. New York: Plenum.
- REHFELDT, R. A., LATIMORE, D., & STROMER, R. (2003). Observational learning and the formation of classes of reading skills by individuals with autism and other developmental disabilities. *Research in Developmental Disabilities*, 25, 333-358.
- REID, D. H., & GREEN, C. W. (1990). Staff training. In J. L. Matson (Ed.), *Handbook of behaviour modification with the mentally retarded* (2nd ed.) (pp. 71-90). New York: Plenum Press.
- SAUNDERS, K. J., & WILLIAMS, D. C. (1998). Stimulus control procedures. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 193-228). New York: Plenum.
- SCHENK, J. J. (1993). Emergent conditional discrimination in children: Matching to compound stimuli. *The Quarterly Journal of Experimental Psychology*, 46B, 345-365.
- SCHENK, J. J. (1994). Emergent relations of equivalence generated by outcome-specific consequences in conditional discrimination. *The Psychological Record*, 44, 537-558.



- SCHENK, J. J. (1995). Complex stimuli in nonreinforced simple discrimination tasks: Emergent simple and conditional discriminations. *The Psychological Record*, 45, 477-494.
- SIDMAN, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, 14, 5-13.
- SIDMAN, M. (1994). *Equivalence relations and behavior: A research story*. Boston, MA: Author's Cooperative.
- SMEETS, P. M., BARNES, D., & ROCHE, B. (1997). Functional equivalence in children: Derived stimulus-response and stimulus-stimulus relations. *Journal of Experimental Child Psychology*, 66, 1-17.
- SMEETS, P. M., & BARNES-HOLMES, D. (2003). Children's emergent preferences for soft drinks: Stimulus equivalence and transfer. *Journal of Economic Psychology*, 24, 603-618.
- SMEETS, P. M., BARNES-HOLMES, Y., AKINPAR, D., & BARNES-HOLMES, D. (2003). Reversal of equivalence relations. *The Psychological Record*, 53, 91-120.
- SMEETS, P. M., DYMOND, S., & BARNES-HOLMES, D. (2000). Instructions, stimulus equivalence, and stimulus sorting: Effects of sequential testing arrangements and a default option. *The Psychological Record*, 50, 339-354.
- SMEETS, P. M., SCHENK, J. J., & BARNES, D. (1995). Establishing arbitrary stimulus classes via identity-matching training and non-reinforced matching with complex stimuli. *The Quarterly Journal of Experimental Psychology*, 48B, 311-328.
- SMEETS, P. M., & STRIEFEL, S. (1994). A revised blocked-trial procedure for establishing arbitrary matching in children. *The Quarterly Journal of Experimental Psychology*, 47B, 241-261.
- STODDARD, L. T. (1982). An investigation of automated methods for teaching severely retarded individuals. In N. R. Ellis (Ed.), *International review of research in mental retardation* (pp. 163-207). New York: Academic Press.
- STROMER, R., MACKAY, H. A., MCVAY, A. A., & FOWLER, T. (1998). Written lists as mediating stimuli in the matching-to-sample performances of individuals with mental retardation. *Journal of Applied Behavior Analysis*, 31, 1-19.
- STROMER, R., MACKAY, H. A., & REMINGTON, R. (1996). Naming, the formation of stimulus classes, and applied behavior analysis. *Journal of Applied Behavior Analysis*, 29, 409-431.
- TIERNEY, K. J., DE LARGY, P., & BRACKEN, M. (1995). Formation of an equivalence class incorporating haptic stimuli. *The Psychological Record*, 45, 431-438.
- WEAVER, A. D., WATSON, T. S., CASHWELL, C., HINDS, J., & FASCIO, S. (2003). The effects of ability- and effort-based praise on task persistence and task performance. *The Behavior Analyst Today*, 4(2), 127-133.
- ZYGMONT, D. M., LAZAR, R. M., DUBE, W. V., & MCILVANE, W. J. (1992). Teaching arbitrary matching via sample stimulus-control shaping to young children and mentally retarded individuals. *Journal of the Experimental Analysis of Behavior*, 57, 109-117.

Copyright of Psychological Record is the property of Psychological Record. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.