

BRIEF REPORT

SUPPLEMENTAL MEASURES OF DERIVED STIMULUS RELATIONS

Simon Dymond and Ruth Anne Rehfeldt

APU, CAMBRIDGE and SOUTHERN ILLINOIS UNIVERSITY

Explanations of behavioral phenomena are often based in part upon the means by which behavior is measured and recorded during experimental sessions. At times, it may be useful to supplement one's research program with novel measures in order to detect procedural dependencies in results, and to prevent one's measures from becoming confused with the behavioral phenomena under study (see L. Hayes, 1992). One area in which this may be the case is in the study of derived stimulus relations, which refers to the emergence of relations that are not explicitly trained, such as equivalence relations (Sidman, 1994) or arbitrarily applicable multiple stimulus relations (see S.C. Hayes, Barnes-Holmes, & Roche, 2001).

The establishment of derived stimulus relations is usually inferred when an individual responds correctly on a certain percentage of test trials assessing the emergence of untrained conditional discriminations. Thus, *percentage of test trials correct* is the measure by which the emergence of derived relations is typically assumed. However, the emergence of derived relations and a test score of 90% accuracy or higher are not one and the same; the latter is merely the most frequently adopted measure for studying the former. It is hence necessary to explore other means of measuring and recording derived stimulus relations, in hopes of shedding new light on the nature of derived responding, as well as to inspire future clinical interventions based upon basic research findings.

In addition to this percentage correct criterion, several types of supplemental measures have been described for use in the study of derived stimulus relations. These additional measures provide convergent validity with respect to the standard measure of derived relations and may also shed light on some of the variables responsible for derived responding. The present paper describes several such measures and highlights ways in which they might be valuable in our attempts to further understand derived phenomena.

Many thanks to Tom Critchfield, Paul Smeets, Bill Dube, and an anonymous reviewer for helpful comments on an earlier version of this article. Address correspondence to Simon Dymond, Department of Psychology, APU, East Road, Cambridge, CB1 1PT (s.dymond@apu.ac.uk).

RESPONSE LATENCY

In repeated exposures to a test for derived stimulus relations, there is a marked transition from inaccurate to accurate responding. While this 'delayed emergence' is itself noteworthy (see Sidman, 1994), it has been argued that only measuring response accuracy may obscure any differences in the time taken to respond to different trial types. Indeed, several recent studies have revealed such differences in the time taken to respond to trained discriminations and probe trials (Bentall, Dickins, & Fox, 1993; Spencer & Chase, 1996; Wulfert & Hayes, 1988). Researchers have measured the reaction time, or response latency, taken to select a comparison stimulus with (Spencer & Chase, 1996) or without (Bentall et al., 1993) an observing response to the sample prior to the simultaneous presentation of the comparison stimuli. Despite the minor differences in procedures (see Spencer and Chase, 1996), research has consistently revealed differences in response latencies to trained baseline and probe trials which covary with other more formal measures of derived stimulus relations such as response accuracy. For example, Wulfert and Hayes (1988) found that subjects' response latencies on baseline and symmetry trials differed significantly from their latencies on transitivity and equivalence trials, even though response accuracy remained the same across all trial types. Spencer and Chase (1996) extended these findings and showed a nodal distance effect for response latency: subjects' latencies of matching equivalence trials separated by one-node were shorter (i.e., faster) than their latencies to five-node equivalence trials.

Recently, O'Hora, Roche, Barnes-Holmes, and Smeets (in press) examined response latencies to the mutually entailed comparison relations of more-than and less-than (see also Steele and Hayes, 1991). O'Hora et al found that response latencies to probes for more-than and less-than relations were significantly greater than latencies to probes for same and opposite relations. Interestingly, this effect was found to decrease significantly following extended training with an exemplar of more-than and less-than relations. The observation by O'Hora et al (in press) that extended exposure to exemplars of more-than/less-than relations leads to a decrease in

response latency highlights the flexible, operant nature of the responding under study - a finding that may not have become apparent were it not for the incorporation of an additional measure of derived stimulus relations.

VERBAL REPORTS

Recording and subsequently analyzing the overt verbal behavior emitted by individuals as they complete equivalence-based experimental tasks is a strategy that has continued to attract the interest of behavior analysts in recent years. Concurrent "think-aloud" procedures (Ericsson & Simon, 1984; S. Hayes, 1986), in which subjects are required to "talk-aloud" everything that they are thinking to themselves over the course of completing an experimental task, have allowed for an investigation of the relationship between overt verbal behavior and the acquisition and emergence of derived relations (Rehfeldt & Dixon, 2000; Rehfeldt, Dixon, Hayes, & Steele, 1998; Wulfert, Dougher, & Greenway, 1991). Results from stimulus equivalence experiments that have employed verbal protocol analyses have raised important questions regarding verbal organisms' learning of stimulus relations and their descriptions of those relations. For example, Wulfert et al. (1991) found that subjects who failed to show class formation were likely to verbally describe sample stimuli and their matching comparisons as unitary stimulus compounds, whereas subjects who did demonstrate class formation frequently described the relations between the matching stimuli.

Similarly, verbal reports, usually collected through postexperimental interviews or questionnaires, have also been used as a supplemental measure of equivalence performance (e.g., Dube, Green, & Serna, 1993). Lane and Critchfield (1996), however, collected verbal reports following every trial of their equivalence experiment. These researchers restricted the range of possible report topographies available by presenting subjects with the computer-generated query "was your selection correct or incorrect?", followed by a confidence rating, "how confident are you about your selection?", after every baseline and probe trial. Subjects' reports generally corresponded to their matching-to-sample (MTS) performances: following a 'correct' MTS trial, subjects usually reported their selections as correct and expressed strong confidence. For half of the subjects in Lane and Critchfield's (1996) study, however, verbal reports on tests for reflexivity described a lower level of selection accuracy than was indicated by their MTS performances.

Another measure which utilizes subjects' verbal reports has been described by Pilgrim and Galizio (1996, pp. 184-186). Collecting what those researchers termed "percentage reinforcement estimates" involved presenting subjects, at the conclusion of an equivalence study, with a series of stimulus pairs that were presented during the training and testing phases. One stimulus of each pair was presented in the sample stimulus position, while the second stimulus was presented in the comparison stimulus position. Participants were then asked to estimate, on a scale from 0 to 100, the percentage of trials over the course of the experiment on which choosing the presented comparison stimulus given the presented sample stimulus was reinforced. Pilgrim and Galizio (1996) reported that subjects' estimates for sample and comparison pairs that had been presented on test trials were highest for symmetry relations, followed by transitivity relations, then reflexivity relations (Pilgrim & Galizio, 1996), despite the fact that all of the test trials were presented in extinction. These reports thus converged with other results reported by the same authors, who found that symmetry relations were most sensitive to reversals in baseline contingencies, relative to transitivity relations (e.g., Pilgrim & Galizio, 1995). Percentage reinforcement estimates thus provided convergent validity with the equivalence test. Findings such as these suggest a useful role for verbal-reports in revealing outcomes not evident in typical measures of equivalence.

STIMULUS SORTING & STIMULUS RECALL

Stimulus sorting tests, although widely used in studies on categorization and concept formation, have only recently been used in studies on derived stimulus relations (Green, 1990; Pilgrim & Galizio, 1996; Smeets, Dymond, & Barnes-Holmes, 2000). During a stimulus sorting task, subjects are presented with all of the individual stimuli and instructed to "place these objects into groups, whatever groups you think are most appropriate" (Pilgrim & Galizio, 1996, p. 188). The evidence available shows that, after demonstrating equivalence, most subjects also sort the stimuli in a class-consistent manner. In a stimulus recall task, subjects are instructed to free recall the equivalence stimuli (Pilgrim & Galizio, 1996, pp.188-190). Based on findings from the cognitive literature showing that stimuli are recalled in clusters based on category membership, Galizio, Stewart, and Pilgrim (in press) demonstrated class clustering in the free recall of equivalence class members. In summary, both stimulus recall and stimulus sorting tasks provide convergent validity with other measures of derived stimulus relations.

Future research employing the additional measures of stimulus sorting and stimulus recall would be well advised to consider both the instructions given to subjects and the presentation format of the task. Inclusion of the instruction "place these objects into two groups" is likely to have a different effect on outcome measures than a instruction to merely categorize the stimuli. Similarly, the format in which the task is presented to subjects is likely to influence the results obtained. For example, Smeets et al., (2000) presented subjects with a sheet containing the instruction to categorize the stimuli into two groups and that "if you find that £ and # belong in one group, put an X next to £ and # in Group 1, and an X next to the other stimuli in Group 2" (p. 346). Class-consistent sorting was shown by 88% of subjects who responded accurately on symmetry and equivalence trials; however, the extent to which this outcome was a function of the detailed instructions and/or presentation format of the task, remains unclear. Stimulus sorting tasks share many characteristics with traditional operant methods used to assess sensitivity to reinforcement contingencies amongst other behaviors (Schlund & Pace, 2000), and thus may provide a cost-effective additional measure of derived stimulus relations for the applied researcher. Stimulus recall measures, which are widely used in cognitive analyses of natural language categories, may be facilitated by the individual features of the stimulus dimensions employed and can provide a useful supplemental measure of equivalence class formation.

STABILITY OVER TIME

Once derived stimulus relations have been measured they are usually not retested at a later date. Saunders, Wachter, and Spradlin (1988), however, demonstrated that equivalence relations may be remarkably stable over time. Four participants with developmental disabilities demonstrated the emergence of derived relations during initial experimental sessions. The four participants were retested for both the directly trained baseline, as well as derived, relations two-to five-months following their original completion of the experiment. Three of the four participants were shown to perform with 90% accuracy or better on tests for both baseline and emergent relations. Similar results were reported by Hollis (1987), who observed test performances ranging from 72% to 87% correct when subjects were tested 100 to 206 days after original training and testing. Thus, as long as the baseline relations are not disrupted, derived relations may be maintained over time in the absence of intervening laboratory experiences (see Spradlin, Saunders, &

Saunders, 1992). Rehfeldt and Hayes (2000) also demonstrated that generalized equivalence relations can be maintained for up to three months in the absence of intervening laboratory experience.

These findings suggest how stability over time may be an important measure to supplement an immediate test for the emergence of derived relations. Assessing the stability over time of derived relations resulting from different training procedures may help researchers evaluate the greater effectiveness of one training procedure relative to another. Examining stability over time may also suggest what types of stimuli might participate in relations that may be better retained, be they arbitrarily configured or naturalistic, or of particular sensory modalities. From a practical perspective, evaluating the period of time over which stimulus relations will persist in the absence of retraining will help to ascertain the value of this teaching approach, and to identify at what point and what amount of retraining is necessary (Dymond & Rehfeldt, 2000).

CONCLUSION

This summary serves to catalog several measures and procedures which have recently been employed in the study of derived stimulus relations. We argued that to avoid confusing the phenomenon under investigation with the measures frequently employed (L. Hayes, 1992), it is important to utilize a wide variety of procedures for evaluating the nature and strength of derived relations. There are several advantages to an incorporation of supplemental measures of derived stimulus relations.

First, relying too closely on one measure such as percentage correct choices may preclude important discoveries regarding the nature of derived stimulus relations. For instance, there is increasing support for the view that deriving stimulus relations is generalized operant behavior and it is likely that in the same way our measurements may obscure exactly what it is we are studying, so too may our experimental procedures, particularly the near-ubiquitous matching-to-sample procedure. Indeed, several authors have argued that the consideration of derived relations other than equivalence and the development of new methods and measures other than those based on matching-to-sample has been restricted by both the explanatory concept of stimulus classes and the idea that equivalence class formation be considered a basic stimulus function (see Barnes-Holmes, Hayes, Dymond, & O'Hora, 2001; Hayes & Barnes, 1997). Second, those working in applied settings may find supplemental measures useful in devising new

and varied instructional programs. Despite the relative dearth of applied interventions based on derived stimulus relations technology, an incorporation of measures other than response accuracy may provide a useful and cost-effective way of facilitating derived responding.

Clearly, supplemental measures will play an important role in informing future research, since applying derived stimulus relations research to an understanding of increasingly complex behavior requires that researchers adapt their measures to the issue under investigation. As researchers continue to extend the scope of their measures to include complex classes of behavioral phenomena, important detail about the conditions necessary and sufficient for the establishment of derived stimulus relations and the usefulness of our empirical analyses for application in the applied arena (Hayes & Hayes, 1993) are likely to result.

REFERENCES

- Barnes-Holmes, D., Hayes, S. C., Dymond, S., & O'Hora, D. (2001). Multiple stimulus relations and the transformation of stimulus functions. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 51-71). New York: Plenum.
- Bentall, R. P., Dickins, D. W., & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *Quarterly Journal of Experimental Psychology*, **46B**, 187-214.
- Dube, W. V., Green, G., & Serna, R. W. (1993). Auditory successive conditional discrimination and auditory stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior*, **59**, 103-114.
- Dymond, S. & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst*, **23**, 239-254.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Galizio, M., Stewart, K., & Pilgrim, C. (in press). Clustering in artificial categories: An equivalence analysis. *Psychonomic Bulletin and Review*.
- Green, G. (1990). Differences in development of visual and auditory-visual equivalence relations. *American Journal on Mental Retardation*, **95**, 260-270.
- Hayes, L. J. (1992). Equivalence as process. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 97-108). Reno, NV: Context Press.
- Hayes, S. C. (1986). The case of the silent dog: Verbal reports and the analysis of rules. A review of K. Anders Ericsson and Herbert A. Simon, "Protocol analysis: Verbal reports as data." *Journal of the Experimental Analysis of Behavior*, **45**, 351-363.
- Hayes, S. C. & Barnes, D. (1997). Analyzing derived stimulus relations requires more than the concept of stimulus class. *Journal of the Experimental Analysis of Behavior*, **68**, 235-244.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York: Plenum.
- Hayes, S. C., & Hayes, L. J. (1993). Applied implications of current JEAB research on derived relations and delayed reinforcement. *Journal of Applied Behavior Analysis*, **26**, 507-511.
- Hollis, J. H. (1987). *Reading vocabulary acquisition and retention in developmentally disabled children*. Working Paper in Child Development, Child Language Program. Lawrence: University of Kansas.
- Lane, S. D. & Critchfield, T. S. (1996). Verbal self-reports of emergent relations in a stimulus equivalence procedure. *Journal of the Experimental Analysis of Behavior*, **65**, 355-374.
- O'Hora, D., Barnes-Holmes, D., & Roche, B. (in press). Response latencies to multiple derived stimulus relations: Testing two predictions of relational frame theory. *The Psychological Record*.
- Pilgrim, C., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior*, **63**, 225-238.
- Pilgrim, C., & Galizio, M. (1996). Stimulus equivalence: A class of correlations or a correlation of classes? In T. R. Zentall and P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 173-195). Amsterdam: Elsevier.
- Rehfeldt, R. & Dixon, M. (2000). Investigating the relation between self-talk and emergent stimulus relations. *Experimental Analysis of Human Behavior Bulletin*, **18**, 28-29.
- Rehfeldt, R. A., Dixon, M. R., Hayes, L. J., & Steele, A. (1998). Stimulus equivalence and the blocking effect. *The Psychological Record*, **48**, 647-664.
- Rehfeldt, R. A., & Hayes, L. J. (2000). The long-term retention of generalized equivalence classes. *The Psychological Record*, **50**, 405-428.

- Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior*, **49**, 95-115.
- Schlund, M. W., & Pace, G. (2000). The effects of traumatic brain injury on reporting and responding to causal relations: An investigation of sensitivity to reinforcement contingencies. *Brain Injury*, **14**, 573-583.
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston, MA: Authors Cooperative.
- Spencer, T. J. & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, **65**, 643-659.
- Spradlin, J. E., Saunders, K. J., & Saunders, R. R. (1992). The stability of equivalence classes. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 29-42). Reno, NV: Context Press.
- 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.
- Steele, D., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior*, **56**, 519-555.
- Wulfert, E., Dougher, M. J., & Greenway, D. E. (1991). Protocol analysis of the correspondence of verbal behavior and equivalence class formation. *Journal of the Experimental Analysis of Behavior*, **56**, 489-504.
- Wulfert, E. & Hayes, S. C. (1988). The transfer of conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior*, **50**, 125-144.