EUROPEAN JOURNAL OF BEHAVIOR ANALYSIS 2012, **13**, 123–135 NUMBER 1 (SUMMER 2012)

Training and Testing Parameters In Formation of Stimulus Equivalence: Methodological Issues

Erik Arntzen
Oslo and Akershus University College

Research on stimulus equivalence has been prominent for more than 40 years in behavior analysis. The present article will focus on issues related to methodological considerations in stimulus equivalence research. An introduction to stimulus equivalence is provided, followed by a discussion of parameters in training conditional discriminations and testing for the emergence of equivalence classes. Some issues related to different training structures, the use of instructions, simultaneous versus delayed matching to sample, the role of familiar stimuli, response requirements for the sample stimulus, and criteria for (a) defining responding in accordance with equivalence and (b) establishing conditional discrimination are discussed in more detail.

Key words: conditional discrimination, stimulus equivalence, parameters, instructions, training structures, delayed matching-to-sample

Since Sidman's first article on stimulus equivalence early in the seventies (Sidman, 1971), interest in stimulus equivalence among researchers has been substantial. Over the last 40 years, there have been a great number of publications on the subject; a search in PsycInfo (June 19, 2012) gave 434 and 288 hits for the terms *stimulus equivalence and functional equivalence*, respectively.

Stimulus equivalence is defined as responding in accordance with the features of reflexivity, symmetry, and transitivity (Sidman, 2000; Sidman & Tailby, 1982). Stimulus equivalence has been demonstrated as a relatively robust phenomenon. Most of the published studies have been concerned with basic research questions (Sidman, 1994), while others have been within the area of applied behavior analysis (e.g., Arntzen, Halstadtro, Bjerke, & Halstadtro, 2010; LeBlanc, Miguel, Cummings, Goldsmith, & Carr, 2003). Furthermore, some of the studies have been concerned with what has lately been characterized as translational research (McIllvane, 2010). However, Sidman was discouraged by the lack of interest in applications: "Puzzled and disillusioned, we soon turned our attention almost exclusively away from applications. Instead, we concentrated our efforts on studying some of the more basic and systematic ramifications of the fascinating phenomenon we had happened upon" (Sidman, 1994, p. 66).

In both basic and applied research on stimulus equivalence formation, the following questions seem to be pertinent: (1) Under what conditions will adult humans not respond in accordance with stimulus equivalence?

Author Note: The author would like to thank three anonymous reviewers and Iver Iversen for their constructive comments and suggestions on an earlier version of the manuscript. Correspondence concerning this manuscript should be addressed to Erik Arntzen, Oslo and Akershus University College, Department for Behavioral Science, PO Box 4 St. Olavs Plass, 0130 Oslo, Norway. E-Mail: erik.arntzen@equivalence.net

(2) What are the most effective parameters for establishing stimulus equivalence? (3) Will the most effective parameters differ across age and participant characteristics?

My interest in stimulus equivalence is fourfold: (1) the emergence of new relations — not directly trained, (2) variables that can influence the emergence of equivalence relations, (3) whether complex repertoires (e.g., concepts, problem solving, language formation) are amenable to behavior analysis, and (4) the impact of research on stimulus equivalence on the arrangement of effective conditional discrimination procedures in behavioral programs. The present article will not go into all of these areas but instead emphasize on one of the most fascinating areas is the emergence of new equivalence relations. Trained relations may be expressed as C x (M-1), and emergent relations could be expressed as C x $(M-1)^2$, where C is the number of classes and M is the number of members. For example, if training potentially three 4-member classes—A1B1C1D1, A2B2C2D2, and A3B3C3D3—there are nine trained relations (A1B1, A2B2, A3B3, B1C1, B2C2 B3C3, C1D1, C2D2, and C3D3) and 27 emergent relations (B1A1, B2A2, B3A3, C1B1, C2B2, C3B3, D1B1, D2B2, D3B3, A1C1, A2C2, A3C3, C1A1, C2A2, C3A3, B1D1, B2D2, B3D3, D1B1, D2B2, D3B3, A1D1, A2D2, A3D3, D1A1, D2A2, and D3A3; see Table 1). The high number of emergent relations when only training a few relations have studied for example by Sidman and colleagues (Sidman, Kirk, & Willson-Morris, 1985).

Much of the research in the area of stimulus equivalence has focused on what happens during testing. As shown in Table 2, a number of variables have been shown to influence successful outcome following conditional discrimination training. It seems that the different results can be attributed to differences in how the training and testing procedures have been arranged.

The purpose of the present article is to discuss some of the variables that influence the outcome of responding in accordance with equivalence (see Table 2). The variables discussed herein are the effects of different training structures, the use of instructions, simultaneous versus delayed matching to sample, the role of familiar stimuli, response requirements for the sample stimulus, and criteria for (a) defining responding in accordance with stimulus equivalence and (b) establishing conditional discrimination. It is important to determine how to produce formation of equivalence classes in the most parsimonious way. Such knowledge will have great importance for arranging an effective technology for teaching. Related to differences in the findings, Sidman (1960) earlier suggested that one should "… inquire experimentally into the reasons for the differences" (p.78). The present article will support this notion and encourage

Table 1. *Classes, Sets of Stimuli, Members, Trained and Emergent Relations*

Classes	Sets of stimuli	Members	Trained Relations		Emergent Relations	
No	Labels	No Labels	C x (M-1) = 3 x (4-1) =9	AB/BC/CD	C x $(M-1)^2$ = 3 x $(4-1)^2$ =27	BA/CB/AC/CA/DC /DB/BD/AD/DA
3	A	4 A1A2A3	Training trials:		Test trials:	
	B	B1B2B3	A1B1 A2B2 A3B3		B1A1 B2A2 B3A3	
	C	C1C2C3	B1C1 B2C2 B3C3		C1B1 C2B2 C3B3	
	D	D1D2D3	C1D1 C2D2 C3D3		D1C1 D2C2 D3C3	
					A1C1 A2C2 A3C3	
					C1A1 C2A2 C3A3	
					B1D1 B2D2 B3D3	
					D1B1 D2B2 D3B3	
					A1D1 A2D2 A3D3	
					D1A1 D2A2 D3A3	

Note. No=number, C=classes, and M=members.

Table 2. *Important Procedural Variables*

Variables and Parameters	Details	Effects
Training protocols	SIM, STC, and CTS	The protocols have shown to have different outcomes.
Training structures	LS, MTO, and OTM	The training structures have different effects on the emergence of equivalence classes.
Number of classes and/or members	Special case is two vs. three/four choices, etc.	It is not clear how an increasing number of members and/or number of classes influences the emergence of equivalence classes depending on training structures.
Requirement of response to sample stimulus or not		This has an effect at least on the number of trials to criterion.
Arrangement of training trials	(a) Gradual introduction of training trials or not (b) Serialized presentation of trials (c) Concurrent presentation of trials	The use of (a), (b), or (c) will have an effect on how fast conditional discriminations are established. Could have an effect on the establishment of equivalence classes.
Arrangement of test trials	(a) Test blocks vs. test trials interspersed in test blocks	Data show that this is probably not critical.
	(b) Equivalence trials presented before symmetry trials	The issue is not clarified.
	(c) Symmetry and equivalence tests or only equivalence tests	It could make a difference. But how, and what kind of difference?
SMTS vs. DMTS		Different processes are going on with SMTS and DMTS.
Intertrial interval		The issue is not clarified.
Response option	Mouse click Keyboard Touch screen	Data show that this is probably not critical
Stimuli used	Abstract stimuli Non-figurative stimuli Nonsense syllabus Pictures	Research has shown that different stimuli could have different effects on the outcome of equivalence classes.
Detailed start-up instruction or not	Such as “belong together”, “are the same”, etc.	This seems to be a critical variable.
Criterion for mastery criterion of conditional discriminations categorizing responding as in accordance with stimulus equivalence or not	A relatively huge variation; 83%–100%	It will have an effect not only on how the classes are established but also on what is defined as responding in accordance with stimulus equivalence or not — an arbitrary boundary.
A phase with changing reinforcement density before testing	Either no phase with reduction of reinforcement density before the tests or a phase with a reduction of probability (i.e., 75%, (50%), 25%, 0%)	This could be important, especially if the arrangement is set up without any reduction of reinforcement before testing.
Characteristics of participants and age	Children Adults Children with autism Dementia patients Elderly participants	It seems like there are differences depending on different training structures.

Note. Different relevant variables and parameters in addition to some details, in the procedures, are shown in the first two columns. Some of the effects are commented on in the far right column. SIM=simultaneous, STC=simple-to-complex, CTS=complex-to-simple, LS=linear series, MTO=many-to-one, and OTM=one-to-many.

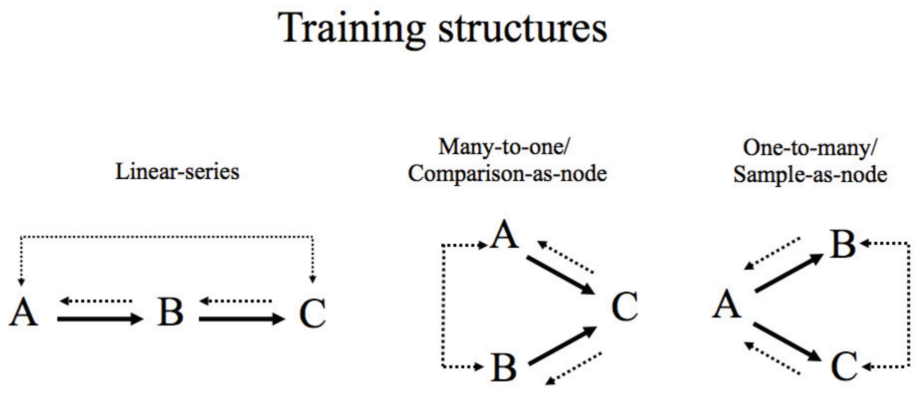

Figure 1. Solid arrows indicate trained relations, dashed-line arrows indicate tests for responding in accord, symmetry, transitivity and equivalence. In this example it is only three members indicating by the letters A, B, and C.

researchers to continue to explore the variables that influence equivalence class formation.

Training Structures

Three training structures have been used in establishing conditional discriminations necessary for testing of emergent relations: linear series (LS), many-to-one (MTO), and one-to-many (OTM) (e.g., K. J. Saunders, Saunders, Williams, & Spradlin, 1993). In potential three 3-member classes, LS involves training of AB and BC relations before testing of emergent relations. MTO involves training of AC and BC relations, and OTM involves training of AB and AC relations (see Figure 1). In the earliest publications on stimulus equivalence, it was assumed, depending on the training structures used to establish the conditional discriminations, that there should not be any differences in equivalence formation among LS, MTO, and OTM training structures (e.g., Sidman, 1994; Sidman & Tailby, 1982). It is difficult to compare equivalence yields across different labs and different studies because the parameters differ. Some studies have found that MTO is the most effective training structure (Fields, Hobbie-Reeve, Adams, & Reeve, 1999; Hove, 2003; K. J. Saunders et al., 1993; R. R. Saunders, Chaney, & Marquis, 2005; R. R. Saunders, Drake, & Spradlin, 1999; R. R. Saunders & McEntee, 2004; R. R. Saunders, Wachter, & Spradlin, 1988; Spradlin & Saunders, 1986), while others have found results in favor of OTM (Arntzen & Holth, 1997, 2000). Furthermore, some studies have found very little difference between MTO and OTM (Arntzen, Grondahl, & Eilifsen, 2010; Arntzen & Hansen, 2011; Arntzen & Nikolaisen, 2011; Arntzen & Vaidya, 2008; Smeets & Barnes-Holmes, 2005). However, the existing results indicate that LS is the least effective training structure in the formation of equivalence classes, at least with a simultaneous protocol, which means that all relations are trained and tested simultaneously (see Imam, 2006, for an extensive description of different protocols). Furthermore, it is important to note that almost all studies that have shown superiority of the MTO training structure have used two classes only. It is therefore a possibility that S-control can explain the differences in stimulus-equivalence outcomes between the studies. Negative contextual control can be reduced by using three or four choices in matching-to-sample training (Carrigan & Sidman, 1992; Johnson & Sidman, 1993; Sidman, 1994). Some other procedural variables that may be responsible for differences in outcomes across studies include the types of test trials (e.g., global equivalence trials only) included in testing (R. R. Saunders et al., 2005), thinning of the probability of programmed consequences, and whether baseline performance was intact during testing (Eilifsen & Arntzen, 2009).

R. R. Saunders and Green (1999) discussed how outcomes on equivalence tests could be related to the number of trained discriminations, and proposed a discrimination analysis that highlighted important differences between the three training structures. Saunders and Green argued that for the participant to respond in accordance with stimulus equivalence, each stimulus must be discriminated from every other stimulus in the experiment. Therefore, outcomes may differ on tests for stimulus equivalence because the number of necessary simple discriminations differ across the training structures. For MTO, every simple discrimination is established, which is not the case for OTM and LS. For example, OTM (e.g., AB and AC) does not require the participant to discriminate between B and C stimuli during baseline training because both are presented as comparisons and, thus, are never presented together. However, during the test phases, participants are required to discriminate between BC and CB stimuli. In a potential three 3-member class after training of AC and BC using MTO, then there will be three discriminations among the C1, C2, and C3 stimuli. Furthermore, C1 will be discriminated from A1, A2, A3, B1, B2, and B3, which will give six discriminations. C2 will be discriminated from A1, A2, A3, B1, B2, and B3, which will give six discriminations. C3 will be discriminated from A1, A2, A3, B1, B2, and B3, which will also give six discriminations. The two samples, A and B, also have to be discriminated because they are mixed in test trials, yielding 15 discriminations (A1A2, A1A3, A2A3, B1B2, B1B3, B2B3, A1B1, A1B2, A1B3, A2B1, A2B2, A2B3, A3B1, A3B2, and A3B3). The total will be 36 discriminations (i.e., 3+6+6+6+15=36).

However, with training of AB and AC (OTM), the number of discriminations among samples will be three. Between comparisons, there will be six discriminations. Furthermore, discriminations between the samples and comparisons will be A1B1, A1B2, A1B3, A2B1, A2B2, A2B3, A3B1, A3B2, A3B3, A1C1, A1C2, A1C3, A2C1, A2C2, A2C3, A3C1, A3C2, and A3C3 (18 discriminations). Thus, OTM entails 27 discriminations (i.e., 3+6+6+18=27). In the LS training structure with three members and three classes, there will be three discriminations between samples and six discriminations between comparisons, along with A1B1, A1B2, A1B3, A2B1, A2B2, A2B3, A3B1, A3B2, A3B3, B1C1, B1C2, B1C3, B2C1, B2C2, B2C3, B3C1, B3C2, and A3B3 (18 discriminations). Thus, LS entails 27 discriminations (i.e., 3+6+18=27). In summary, with three 3-member classes, the number of discriminations required in MTO is 36, versus 27 in OTM and LS. So there are nine discriminations that are not presented in OTM and LS. This difference will increase with the number of members in a class and/or the number of classes. Relatively few experiments, however, have been conducted to test whether the difference in the outcomes on equivalence tests following different training structures changes as a function of increasing class members and/or number of classes (Arntzen, Grondahl, et al., 2010; Arntzen & Hansen, 2011; Fields et al., 1999; R. R. Saunders et al., 2005). For example, in Arntzen and Hansen's study, in which potentially three 6-member equivalence classes were trained, there were only small differences in formation of stimulus equivalence after OTM and MTO training. This could indicate that other variables are as important as the number of simple discriminations during training.

One other possibility that could account for the different outcomes between MTO, LS, and OTM is related to the number of successive and simultaneous discriminations (Arntzen, 2011; Sidman, 2011).

In the following an example with three members and only one class is used for illustration purposes only. Because many samples (A1 and B1) are trained to one comparison (C1) in MTO, while one sample (A1) is trained to many comparisons (B1 and C1) in OTM, there is a difference between the number of successive discriminations of sample stimuli (two times higher in MTO than in OTM) and simultaneous discriminations of comparison stimuli (two times higher in OTM than in MTO), depending on the training structure. In addition, in the test for emergent relations for MTO, some of the stimuli that have served as samples in training will serve as comparisons in the test because the participants have discriminate every sample successively and the comparisons simultaneously. For OTM, some of the stimuli that have served as comparisons during training will serve as samples during testing. For LS (AB and BC) there will be a special case in which the B's will serve as both comparisons and samples. Further research needs to be done to elaborate the effects of the change from successive to simultaneous discriminations and, inversely, the change from simultaneous to successive discriminations. In addition, the effects of stimuli serving both as samples and comparisons seem to be an interesting research question.

Instructions

Sidman (1992) pointed out that one should be careful about instructions in experiments on stimulus equivalence: "Until we have answered the question of whether rules give rise to equivalence, or equivalence makes rules possible, we are going to have to be careful about our experimental procedures in investigations of equivalence. If we tell our subjects that stimuli 'go with' each other (or that they 'match each other,' 'belong together,' 'are the same,' 'go first' or 'go second,' etc.), the data may then tell more about the subject's verbal history than about the effects of current experimental operations" (pp. 21–22).

A number of studies have shown that instructions can play an important role in the development of conditional discriminations (Devany, Hayes, & Nelson, 1986; Pilgrim, Jackson, & Galizio, 2000; Schilmoeller, Schilmoeller, Etzel, & LeBlanc, 1979; Zygmont, Lazar, Dube, & McIlvane, 1992). Hence, it seems important to ask questions about the role of instructions in the formation of equivalence classes. Pilgrim et al. (2000) found that instructions or instructions plus sample naming increased the establishment of arbitrary conditional discriminations in 3- to 6-year-old children. Arntzen, Vaidya, and Halstadtro (2008) raised a number of questions about the effectiveness of the instructions. One issue is timing: It could be that instructions given after participants have failed to respond correctly could be more effective than instructions given at the very beginning of training. "For example, instructions presented at the beginning, when the participant is unfamiliar with many aspects of the task, may be less effective than instructions delivered after the participant has become familiar with the basic requirements of the task, such as orienting to stimuli, scanning the stimulus array, using a mouse or another device to select and respond to a stimulus, and so forth" (Arntzen et al., 2008, p. 29). Another variable is the content of the instruction. It could be that more general instructions like "belong together" are more effective in establishing a rule that ensures that the participant will respond more effectively with novel exemplars. Therefore, Arntzen et al. wanted to study the effects of general instructions presented by the experimenter after the participants did not pass the training criterion for the first conditional discrimination (AB). The participants had approximately 580 trials without any progress in the establishment of the conditional discriminations. The results showed that the instruction "belong together" was effective in establishing conditional discrimination performance of the AB relations. Furthermore, additional relations were established without any further training. All nine participants reached the mastery criterion for the conditional discrimination training. All participants passed the tests for symmetry, and two of nine passed for equivalence.

There are some limitations with the Arntzen et al. (2008) study. For example, we cannot be sure that the presentation of instructions actually influenced the establishment of AB conditional discrimination. Some researchers have argued for a type of learning-set outcome in arbitrary matching that is based on the empirical observation that novel successive arbitrary matching relations are faster for the first relation compared to the next trained relations (K. J. Saunders & Spradlin, 1990, 1993). However, the number of trials for the establishment of AB conditional discriminations in the present study was about 8 times greater than the number of trials for the establishment of AC conditional discriminations, which were trained after the presentation of the instruction. Participants' performance prior to the instruction was poorer than their performance following the instruction. Therefore, the data suggest that the instructions played an important role in establishing conditional discriminations in children in the present study. Future studies should more systematically investigate (1) how the content of the instructions could influence the outcome on equivalence tests and (2) younger participants with varying levels of verbal behavior.

Requirement of a Response to the Sample Stimulus

In the non-human literature a number of studies have been done regarding the importance of observing responses in conditional discrimination procedures (Eckerman, Lanson, & Cumming, 1968; Spetch & Treit, 1986). Eckerman et al. (1968) showed that accuracy increased with requirement of response to sample stimulus and Spetch and Treit (1986) found that number responses to the sample stimulus increased accuracy and they argued that requirement of a high number of responses to sample stimuli ensured exposure to the sample. In a study with college students as participants, Carlin, Wirth, and Chase (1998) compared conditions with and without a requirement of a response to the sample stimulus and they found that a response to sample stimulus increased accuracy. Most of the research involving conditional discrimination procedures has included a requirement of a response to the sample stimulus. The arrangement of conditional discriminations trials could differ as follows: (a) A sample is presented followed by one or more responses to the sample and subsequent presentation of a certain number of comparisons, (b) a sample and a certain number of comparison stimuli are presented at the same time, or (c) a sample is presented and a certain number of comparison stimuli are presented after an interval (usually 1–2 s). The author is not aware of any studies that have compared the effects of different procedural arrangements of sample–comparison presentation. However, in our lab, we have studied the effect of alternatives (a) and (b) (Arntzen, Braaten, Lian, & Eilifsen, 2011), and results showed a faster training for alternative (a) than for alternative (b). These findings could for the role of observing behavior. There is a need for a comparison of all three alternatives in future research.

Pictures or Meaningful Stimuli

Several studies have shown that different types of stimuli may affect the formation of stimulus equivalence. Some types of stimuli may facilitate the formation of stimulus equivalence, while others may hinder equivalence class formation. For example, an increase in formation of stimulus equivalence has been shown with one or more sets of stimuli which have been nameable (Dickins, Bentall, & Smith, 1993), pictures (Arntzen, 2004; Arntzen & Lian, 2010; Arntzen & Nikolaisen, 2011; Holth & Arntzen, 1998), or acquired discriminative function

(Fields, Arntzen, Nartey, & Eilifsen, 2012; Tyndall, Roche, & James, 2004). In other studies, it has been shown that some types of stimuli do not easily become part of an equivalence class (e.g., McGlinchey & Keenan, 1997). McGlinchey and Keenan trained potentially two 3-member classes in 28 primary school children randomly selected from one Catholic and one Protestant school, 2 older children, and 6 adults. They used an LS training structure in which A and C stimuli were Protestant or Catholic stimuli and B stimuli were arbitrary stimuli. In the presence of A1 (Protestant stimulus), selecting B1 (arbitrary stimulus) was correct, while in the presence of A2 (Catholic stimulus), B2 was correct. In addition, in the presence of B1, C1 (Catholic stimulus) was correct, and in the presence of B2, selecting C2 (Protestant stimulus) was correct. They also included arbitrary and nonarbitrary novel stimuli (Protestant and Catholic names) in the tests for emergent relations. They found that in 36% of the cases, the participants did not respond in accordance with stimulus equivalence.

In training of potential classes, the position of the picture stimulus has been studied. The main findings are that familiar pictures or familiar presented as nodes give the highest equivalence yields (e.g., Arntzen & Lian, 2010). For example, if the conditional discriminations have been trained in an LS structure (with three potential 3-member classes; AB and BC relations), having the pictures as B-stimuli has been the most effective arrangement (Holth & Arntzen, 1998). Furthermore, to have the pictures as A stimuli is more effective than presenting them later in training as C stimuli (Arntzen, 2004).

Most of the studies that use familiar stimuli have used conditional discrimination training on a serialized basis, which means that AB trials are trained first, followed by BC trials, then followed by mixing of AB and BC trials. Another way to arrange the training is training on a concurrent basis, presenting a mix of AB and BC trials from the beginning. In an arrangement with concurrent presentation of baseline trials, the order of presentation should not be important. Future research should focus not only on the importance of how the training trials are introduced in general but also on how the introduction of training trials may interact with familiar or unfamiliar stimuli.

Simultaneous Matching to Sample Versus Delayed Matching to Sample

Conditional discrimination procedures used in stimulus equivalence research have, for the most part, been arranged as simultaneous matching to sample (SMTS). In SMTS, a response to the sample is followed by the presentation of comparisons. Both sample and comparisons are presented on the screen until one of the comparisons is chosen. However, some studies have used delayed matching to sample (DMTS; (Arntzen, 2006; Arntzen, Galaen, & Halvorsen, 2007; Arntzen & Haugland, 2012; Bortoloti & De Rose, 2012; R. R. Saunders et al., 2005; Tomanari, Sidman, Rubio, & Dube, 2006; Vaidya & Smith, 2006). In DMTS, the offset of the sample is followed by a delay until the onset of the comparisons. In these delayed matching-to-sample experiments, either a 0-s delay or an *n*-s delay between the offset of the sample and onset of the comparisons has been used, as either one delay value or as a comparison of different delay values. In the studies that have compared different delay values, the main findings have been that longer values have resulted in high accuracy during training (DeFulio, 2002) and have resulted in high equivalence yields (Arntzen, 2006).

It is reasonable to assume that different behavioral processes are going on in SMTS as compared with DMTS procedures. As Sidman (1969) has pointed out, some type of behavior needs to fill the gap between the sample offset and the comparison onset. Blough (1959) showed that for two of four pigeons some stereotyped behavior shown in the delay between the sample offset and the comparison onset could have function as a type of mediation behavior which controlled the choice behavior. Related to humans, the question is, what type of behavior could this be? Furthermore, what is the status of that type of behavior? Some research has found that when distracters are presented during testing, the number of participants responding in accordance with stimulus equivalence has been reduced (Arntzen, 2006). This could indicate that some rehearsal behavior is going on in the gap between the sample offset and the comparison onset, and that distracters may hinder the rehearsal behavior.

Criterion for Correct Responding During Training and Test

A low mastery criterion in training could be troublesome, especially when the conditional discrimination training is arranged to establish potentially only two classes of stimuli. Another issue is the number of training trials per block in conditional discrimination training. If the conditional discrimination training is arranged to establish three potential 3-member classes, each training block consists of 18 trials, and the mastery criterion is, for example, 88 %, then each trial type or sample–comparison is presented three times. In theory, it could be that for one trial type, two of three trials are incorrect and still the participant will respond in accordance with the mastery criterion if he or she has 16 of 18 trials correct. A recommendation would be to have a high mastery criterion for training; 95 % or higher.

The criterion for defining responding in accordance with stimulus equivalence is arbitrary. The criterion used differs across experiments within stimulus equivalence research. However, customarily a criterion of responding in accordance with stimulus equivalence is at 90% or above. It could be that some of the differences in outcome on equivalence tests across laboratories are related to differences in mastery criterion. We should be careful with the criterion for the definition of responding in accordance with stimulus equivalence, especially when using two classes.

Conclusion

The present article has focused on some methodological variables or parameters in stimulus equivalence research that could influence the outcome on the test for the emergence of equivalence classes. Some of these variables have been discussed. For example, the experiment with instructions is a demonstration of how a certain type of instruction can influence the establishment of conditional discrimination. Furthermore, the present article raises some more general questions, for example, can differences in stimulus equivalence outcome be related to the different instructions employed? Another issue that seems to be of substantial importance is the criteria used to document adequate training and the criteria used in test for categorizing responding in accordance with stimulus equivalence. For example, mastery criterion should be higher in training (e.g., 95–100%) than in testing (e.g., 90–100%). Finally, the evaluation of different variables will hopefully facilitate the development of an effective teaching of stimulus equivalence classes.

References

Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record, 54*, 275–291. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Arntzen, E. (2006). Delayed matching to sample and stimulus equivalence: Probability of responding in accord with equivalence as a function of different delays. *The Psychological Record, 56*, 135–167. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Arntzen, E. (2011). Comments on Sidman's Remarks. *European Journal of Behavior Analysis, 12*, 351–353. Retrieved from http://www.ejoba.org/.

Arntzen, E., Braaten, L. F., Lian, T., & Eilifsen, C. (2011). Response-to-sample requirements in conditional discrimination procedures. *European Journal of Behavior Analysis, 12*, 505–522. Retrieved from Retrieved from http://www.ejoba.org/.

Arntzen, E., Galaen, T., & Halvorsen, L. R. (2007). Different retention intervals in delayed matching-to-sample: Effects of responding in accord with equivalence. *European Journal of Behavior Analysis, 8*, 177–191. Retrieved from http://www.ejoba.org/.

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and the subsequent performance on the tests for stimulus equivalence. *The Psychological Record, 60*, 437–462. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Arntzen, E., Halstadtro, L. B., Bjerke, E., & Halstadtro, M. (2010). Training and testing theoretical music skills in a boy with autism using a matching-to-sample format. *Behavioral Interventions*, 129–143. doi: 10.1002/bin.301

Arntzen, E., & Hansen, S. (2011). Training Structures and the Formation of Equivalence Classes. *European Journal of Behavior Analysis, 12*, 483–503. Retrieved from http://www.ejoba.org/.

Arntzen, E., & Haugland, S. (2012). Titration of limited hold to comparison in conditional discrimination. *The Psychological Record, 62*, 243–262. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record, 47*, 309–320. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Arntzen, E., & Holth, P. (2000). Differential probabilities of equivalence outcome in individual subjects as a function of training structure. *The Psychological Record, 50*, 603–628. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Arntzen, E., & Lian, T. (2010). Trained and derived relations with pictures as nodes. *The Psychological Record, 60*, 659–677. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Arntzen, E., & Nikolaisen, S. L. (2011). Establishing equivalence classes in children using familiar and abstract stimuli and many-to-one and one-to-many training structures. *European Journal of Behavior Analysis, 12*, 105–120. Retrieved from http://www.ejoba.org/.

Arntzen, E., & Vaidya, M. (2008). The effect of baseline training structure on equivalence class formation in children. *Experimental Analysis of Human Behavior Bulletin, 29*, 1–8. Retrieved from http://www.eahb.org/NewSitePages/BulletinHomepage.htm.

Arntzen, E., Vaidya, M., & Halstadtro, L. B. (2008). On the role of instruction in conditional discrimination training. *Experimental Analysis of Human Behavior Bulletin, 29*, 17–24. Retrieved from http://www.eahb.org/NewSitePages/BulletinHomepage.htm.

Blough, D. S. (1959). Delayed matching in the pigeon. *Journal of the Experimental Analysis of Behavior, 2*, 151–160. doi: 10.1901/jeab.1959.2-151

Bortoloti, R., & De Rose, J. C. (2012). Equivalent stimuli are more strongly related after training with delayed matching than after simultaneous matching: A study using the implicit relational assement procedure (IRAP). *The Psychological Record, 62*, 41–54.

Carlin, L. A., Wirth, O., & Chase, P. N. (1998). Effects of sample response requirements on matching-to-sample performance with humans. *Experimental Analysis of Human Behavior Bulletin, 16*, 2–5. Retrieved from http://www.eahb.

org/NewSitePages/BulletinHomepage.htm.

Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 58*, 183–204. doi: 10.1901/jeab.1992.58-183

DeFulio, A. L. (2002). *A comparison of auditory and visual stimuli in a delayed matching to sample procedure with adult humans.* Unpublished manuscript, University of North Texas, Denton. Retrieved from http://www.library.unt.edu/theses/open/20023/defulio_anthony/thesis.pdf

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior, 46*, 243–257. doi: 10.1901/jeab.1986.46-243

Dickins, D. W., Bentall, R. P., & Smith, A. B. (1993). The role of individual stimulus names in the emergence of equivalence relations: The effects of interpolated paired-associates training of discordant associates between names. *The Psychological Record, 43*, 713–724.

Eckerman, D. A., Lanson, R. N., & Cumming, W. W. (1968). Acquisition and maintenance of matching without a required observing response. *Journal of the Experimental Analysis of Behavior, 11*, 435–441. doi: 10.1901/jeab.1968.11-435

Eilifsen, C., & Arntzen, E. (2009). On the role of trial types in tests for stimulus equivalence. *European Journal of Behavior Analysis, 10*, 187–202. Retrieved from http://www.ejoba.org/.

Fields, L., Arntzen, E., Nartey, R., & Eilifsen, C. (2012). Effects of a meaningful, a discriminative, and a meaningless stimulus on equivalence class formation. *Journal of the Experimental Analysis of Behavior, 97*, 163–181. doi: 10.1901/jeab.2012.97-163

Fields, L., Hobbie-Reeve, S. A., Adams, B. J., & Reeve, K. F. (1999). Effects of training directionality and class size on equivalence class formation by adults. *The Psychological Record, 49*, 703–724. Retrieved from http://thepsychologicalrecord.siuc.edu/.

Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record, 48*, 81–110. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Hove, O. (2003). Differential probability of equivalence class formation following a one-to-many versus a many-to-one training structure. *The Psychological Record, 53*, 617–634. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior, 85*, 107–124. doi: 10.1901/jeab.2006.58-04

Johnson, C., & Sidman, M. (1993). Conditional discrimination and equivalence relations: Control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 59*, 333–347. doi: 10.1901/jeab.1993.59-333

LeBlanc, L. A., Miguel, C. F., Cummings, A. R., Goldsmith, T. R., & Carr, J. E. (2003). The effects of three stimulus-equivalence testing conditions on emergent US geography relations of children diagnosed with autism. *Behavioral Interventions, 18*, 279–289. doi: 10.1002/bin.144

McGlinchey, A., & Keenan, M. (1997). Stimulus equivalence and social categorization in Northern Ireland. *Behavior and Social Issues, 7*, 113–128.

McIllvane, W. J. (2010). Translational behavior analysis: From laboratory science in stimulus control to intervention with persons with neurodevelopmental disabilities. *The Behavior Analyst, 32*, 273–280.

Pilgrim, C., Jackson, J., & Galizio, M.

(2000). Acquisition of arbitrarily conditional discriminations by young normally developing children. *Journal of the Experimental Analysis of Behavior, 73*, 177–193. doi: 10.1901/jeab.2000.73-177

Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record, 43*, 725–744. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Saunders, K. J., & Spradlin, J. E. (1990). Conditional discrimination in mentally-retarded adults: The development of generalized skills. *Journal of the Experimental Analysis of Behavior, 54*, 239–250. doi: 10.1901/jeab.1990.54-239 Retrieved from ISI:A1990EF96500009

Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: Programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior, 60*, 571–585. doi: 10.1901/jeab.1993.60-571

Saunders, R. R., Chaney, L., & Marquis, J. G. (2005). Equivalence class establishment with two-, three-, and four-choice matching to sample by senior citizens. *The Psychological Record, 55*, 539–559. Retrieved from http://thepsychologicalrecord.siuc.edu/.

Saunders, R. R., Drake, K. M., & Spradlin, J. E. (1999). Equivalence class establishment, expansion, and modification in preschool children. *Journal of the Experimental Analysis of Behavior, 71*, 195–214. doi: 10.1901/jeab.1999.71-195

Saunders, R. R., & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior, 72*, 117–137. doi: 10.1901/jeab.1999.72-117

Saunders, R. R., & McEntee, J. E. (2004). Increasing the probability of stimulus equivalence with adults with mild mental retardation. *The Psychological Record, 54*, 423–435. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Saunders, R. R., Wachter, J. A., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedure. *Journal of the Experimental Analysis of Behavior, 49*, 95–115. doi: 10.1901/jeab.1988.49-95

Schilmoeller, G. L., Schilmoeller, K. J., Etzel, B. C., & LeBlanc, J. M. (1979). Conditional Discrimination After Errorless and Trial-And-Error Training. *Journal of the Experimental Analysis of Behavior, 31*, 405–420. doi: 10.1901/jeab.1979.31-405

Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.

Sidman, M. (1969). Generalization gradients and stimulus control in delayed matching-to-sample. *Journal of the Experimental Analysis of Behavior, 12*, 745–757. doi: 10.1901/jeab.1969.12-745

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 14*, 5–13. Retrieved from http://jslhr.asha.org/.

Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding Verbal Relations* (pp. 15–27). Reno, Nevada: Context Press.

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

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*, 127–146. doi: 10.1901/jeab.2000.74-127

Sidman, M. (2011). Reply to commentaries on "Remarks" columns. *European Journal of Behavior Analysis, 12*, 355–370. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six members stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior, 43*, 21–42. doi: 10.1901/jeab.1985.43-21

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5–22. doi: 10.1901/jeab.1982.37-5

Smeets, P. M., & Barnes-Holmes, D. (2005). Establishing equivalence classes in preschool children with one-to-many and many-to-one training protocols. *Behavioural Processes, 69*, 281–293. doi: 10.1016/j.beproc.2004.12.009

Spetch, M. L., & Treit, D. (1986). Does effort play a role in the effect of response requirements on delayed matching to sample? *Journal of the Experimental Analysis of Behavior, 45*, 19–31. doi: 10.1901/jeab.1986.45-19

Spradlin, J. E., & Saunders, R. R. (1986). The development of stimulus classes using match-to-sample procedures: Sample classification versus comparison classification. *Analysis and Intervention in Developmental Disabilities, 6*, 41–58. doi: 10.1016/0270-4684(86)90005-4

Tomanari, G. Y., Sidman, M., Rubio, A. R., & Dube, W. V. (2006). Equivalence classes with requirements for short response latencies. *Journal of the Experimental Analysis of Behavior, 85*, 349–369. doi: 10.1901/jeab.2006.107-04

Tyndall, I. T., Roche, B., & James, J. E. (2004). The relation between stimulus function and equivalence class formation. *Journal of the Experimental Analysis of Behavior, 81*, 257–266. Retrieved from ISI:000223182300003

Vaidya, M., & Smith, K. N. (2006). Brief report: Delayed matching-to-sample training facilitates derived relational responding. *Experimental Analysis of Human Behavior Bulletin, 24*, 9–16. Retrieved from http://www.eahb.org/NewSitePages/BulletinHomepage.htm.

Zygmont, D., Lazar, R., 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. doi: 10.1901/jeab.1992.57-109