John Carroll University
Carroll Collected

Psychology

9-2013

Test order effects in simultaneous protocols

Abdulrazaq Imam
John Carroll University, aimam@jcu.edu

Timothy A. Warner

Follow this and additional works at: http://collected.jcu.edu/psyc-facpub

Part of the Psychology Commons

Recommended Citation

Test order effects in simultaneous protocols

Abdulrazaq A. Imam & Timothy A. Warner

Learning & Behavior

VOLUME 41, NUMBER 3 • SEPTEMBER 2013

ONLINE FIRST

L&B

Springer

Springer

Springer

Test order effects in simultaneous protocols

Abdulrazaq A. Imam · Timothy A. Warner

Abstract Simultaneous protocols typically yield poorer stimulus equivalence outcomes than do other protocols commonly used in equivalence research. Two independent groups of three 3-member equivalence sets of stimuli were used in conditional discrimination procedures in two conditions, one using the standard simultaneous protocol and the other using a hybrid simultaneous training and simple-to-complex testing. Participants completed the two conditions in one long session in Experiment 1, but in separate sessions in Experiment 2. The same stimulus sets used in Experiment 1 were randomized for the two conditions in Experiment 2. Overall, accuracy was better with the hybrid than with the standard protocol in both experiments. The equivalence yield was also better under the hybrid than under the standard protocol in each experiment. The results suggest that the order of testing for emergent relations may account for the difficulty often encountered with the standard simultaneous protocol.

Keywords Stimulus equivalence · Matching to sample · Simultaneous protocol · Simple-to-complex protocol · Touch screen · Humans

In the past three decades or so, behavioral psychology has witnessed an explosion of interest in stimulus equivalence, following initial reports by Sidman and Tailby (1982) on the subject. This line of research has fostered a behavioral account of complex phenomena hitherto considered beyond the purview of a behavioral analysis. Many advances have been made methodologically among the many successes of the paradigm (see Sidman, 1994). One such advance is the observation of differential equivalence outcomes based on procedural configurations of the stimulus presentations during training and testing of conditional relations among the potential participating stimuli. Two such configurations have proven critically important in their effects on stimulus equivalence outcomes: training structures (Saunders & Green, 1999) and protocols (Fields et al., 1997). Ordinarily, given stimulus sets A, B, and C, Sidman equivalence requires that after conditional discrimination training of AB and BC relations, participants should be able to select the A and B comparison stimuli given B and C samples (BA and CB symmetry), and to select C and A comparisons given A and C samples, respectively (AC transitivity and CA equivalence), to establish that these stimuli belong in an equivalence class. Training AB and then BC in this example defines a linear-series (LS) training structure, which has proven generally to make establishing equivalence more difficult than do the other two structures—comparison-as-node (CAN) or many-to-one (MTO), defined by BA and CA training, and sample-as-node (SAN) or one-to-many (OTM), defined by AB and AC training (see Saunders & Green, 1999) (some controversy has attended the relative effectiveness of the latter two structures, however; see, e.g., Arntzen & Holth, 2004; Hove, 2003).

It turns out that demonstrating equivalence also depends on whether the conditional discriminations were trained and tested all at once, one at a time, or in what particular order. Referred to as *protocols*, three such different procedural configurations—simple-to-complex (STC), complex-to-simple (CTS), and simultaneous (SIM)—have been identified and used extensively in the literature (Fields et al., 1997). In the STC protocol, individual relations first are taught separately (e.g., BA, then CA) and then together (e.g., BA and CA in one block), interspersed with tests for derived relations separately (e.g., AB, then AC) and then together (AB and AC in one block), before a final test for equivalence (BC and CB)

A. A. Imam · T. A. Warner
John Carroll University, University Heights, OH, USA

A. A. Imam (✉)
Department of Psychology, John Carroll University, 1 John Carroll Blvd., University Heights, OH 44118, USA
e-mail: aimam@jcu.edu

intermixed with baseline trials. In a CTS protocol, training of each baseline relation (e.g., BA and CA) would be followed by a combined test for equivalence (e.g., BC and CB) before the other derived relations (e.g., AB and AC) are then tested (see Imam, 2006, for an LS example). In contrast to these protocols, the SIM protocol implements training of all baseline relations (e.g., BA, CA, etc.) in one block before testing for all emergent relations (e.g., AB, CA, BC, CB, etc.) in another block. The main difference between the STC and SIM protocols thus is that the former trains and tests relations incrementally and interlaced, whereas the latter trains all relations together before testing for any derived symmetry, transitivity, or equivalence relations. A number of studies have shown major differences in the equivalence outcomes based on these protocols (Adams, Fields, & Verhave, 1993; Fields et al., 1997). The STC tends to yield better results than do the other two protocols (i.e., CTS and SIM), whereas the SIM protocol tends to fare the worst: Fewer participants tend to form equivalence classes, unless they have had pretraining with the STC protocol (Buffington, Fields, & Adams, 1997; Fields, Landon-Jimenez, Buffington, & Adams, 1995; Fields et al., 1997). The relatively limited success of the SIM protocol has remained unexplored.

What is it about the SIM protocol particularly that predisposes it to failure more than the other protocols? The obvious candidate for comparison is the STC protocol, because of its enhanced level of success. A unique feature of this protocol, in contrast to the SIM protocol, is that it intersperses the requisite baseline relations with simpler to more complex testing, in a linear fashion (see Imam, 2006). For example, BA training would require AB symmetry testing, before introducing CA training, followed by AC symmetry testing, before the BC and CB equivalence tests, and so on. In contrast, the SIM protocol would involve training of AB and AC relations together before testing for the BA, CA symmetry and the BC, CB equivalence all together. The sequential, linear training of the individual baseline relations, along with the interlacing of each with simpler to more advanced tests, appears to be critical in the better yield that the STC affords. This is in contrast to the massed nature of training and then testing in the SIM protocol, and the relative contributions of the training sequencing versus the simple-to-more-complex testing have not been examined empirically. In showing that higher-order relations that previously had failed to emerge could be demonstrated following testing of lower-order relations, Sidman, Kirk, and Willson-Morris (1985) suggested that the better equivalence outcome in STC could be due to the verification of lower-order derived relations before higher-order relations in that protocol, a procedure not afforded by the SIM protocol.

What, then, would happen if, following SIM training, the tested relations, rather than being combined in a single test as usual, were dispersed and sequenced to simulate the testing in STC? The purpose of the present study was to examine whether such a hybrid of simultaneous training and simple-to-complex testing would yield better test performance than the standard SIM protocol. To minimize the relative impact of training structure, the CAN (or MTO; Hove, 2003; Saunders & Green, 1999) structure was adopted. If the testing features of the STC protocol are a critical factor in its success, the hybrid protocol in the present study should yield better equivalence outcomes than the standard protocol.

Experiment 1

Method

Participants

The participants consisted of 16 undergraduate students—ten males and six females. The students were primarily freshman and sophomores, in addition to one senior. Participants' ages were between 18 and 23 years old, and all students were enrolled at John Carroll University. All of the participants had no previous research experience.

Apparatus and stimuli

The MTS software (Dube & Hiris, 1997) programmed the experimental events and collected data on a Macintosh computer. All stimuli appeared on a white square (4.7×4.7 cm) against a black background, with the sample stimulus always being displayed at the center of the screen and the three comparison stimuli at the corners of the screen. The locations of the comparison stimuli were determined randomly from trial to trial, always leaving a corner blank. Figure 1 presents the 2.5×2.5 cm stimuli used for the hybrid protocol (top) and for the standard simultaneous protocol (bottom). Participants did not know the letter and number designations of, respectively, the comparison stimuli and the class memberships.

Procedures

Computer touchscreen responses registered stimuli selections and automatically recorded response latencies. The computer also determined consequences for each response, and recorded class-consistent responses as correct and other responses as incorrect.

Pretraining A demonstration of the MTS procedure was conducted for one correct and one incorrect response, using upper- and lowercase English letters as the sample and comparison stimuli, respectively. Participants then completed 24 trials using the remaining English letters. No special instructions accompanied the demonstration. We assumed that
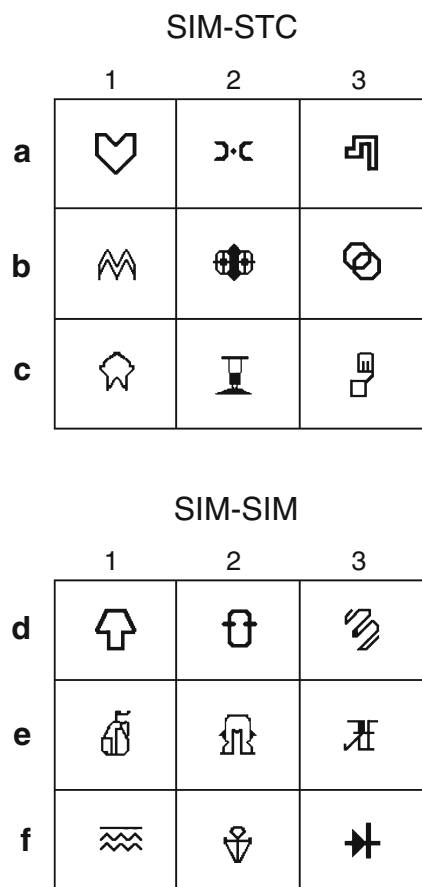
SIM-STC

SIM-SIM

Fig. 1 Stimuli used in the hybrid SIM–STC (top) and the standard SIM–SIM (bottom) protocols in Experiment 1. The letters designate sets of sample and comparison stimuli, and the numbers designate potential stimulus classes

participants, as college students, had a repertoire of identity matching, and therefore no test of reflexivity was conducted.

Matching to sample Training and testing occurred in separate blocks of trials in a single session. A trial started with a sample stimulus appearing at the center of the screen. Touching the sample stimulus produced three comparison stimuli in the corners of the screen. Selecting the correct comparison on trials with feedback produced a 1-s display of the word "correct" and a tone, incremented a hidden counter, and started a 1.5-s intertrial interval (ITI). Selecting an incorrect comparison ended the trial, sounded a buzzer while also darkening the screen for 1 s, and began the ITI. Touching on the blank stimulus location was considered incorrect, and a touch during the ITI reset the ITI timer, ensuring that 1.5 s had elapsed without a response.

Training involved three levels of feedback (100 %, 50 %, and 0 %), with each level starting training blocks with varying levels of instruction. The following instruction appeared for the 100 %-feedback blocks: "You will receive feedback for the next block of trials. In addition to 'Correct', you will hear a tone for a correct choice, and you will hear a buzzer in

addition to a black screen for an incorrect choice." With the 50 % feedback, each block started with "You will receive feedback for ONLY SOME trials in the next block of trials." At the beginning of the 0 %-feedback and testing blocks, participants were told "In the next block of trials, you will be given no feedback regarding the accuracy of your responses. 'Correct' will not be displayed, you will hear no tone when you make a correct choice; you will hear no buzzer, and the screen will not turn black when you make an incorrect choice."

Training and testing Two sets of conditional relations (B1A1, B2A2, B3A3; C1A1, C2A2, C3A3), representing the CAN or MTO training structure, were trained in each condition. The sequence of training and testing trials within a block was randomly determined. Each training block, regardless of the feedback level or condition, consisted of 12 trials, as is shown in Table 1 (top panel), for each protocol. Participants experienced the feedback levels in the same order (100 %, 50 %, and 0 %) and began testing after the 0 %-feedback trials. The criterion for advancing through the feedback levels to testing under the standard and hybrid protocols was at least 90 % correct. Incorrect responses on five trials in the 12-trial training blocks would automatically initiate retraining with 100 % feedback, but this was not required in the present experiment. Accuracy of at least 90 % in the equivalence testing was required under the hybrid SIM–STC condition and on each mixed-test block under the standard SIM–SIM condition, with no more than five blocks being completed during these tests.

Protocols Participants received no instructions describing the differences between the two protocols. Training involved one set of sample–comparison stimuli at a time under each of the protocols, respectively, depicted in Fig. 1. In the standard SIM training used in both protocols, all six baseline (BA, CA) trial types appeared twice randomly in a single training block, two versions of which alternated until participants achieved performance criterion. In testing blocks under the SIM–SIM protocol, all relations, including the baseline (BA, CA), symmetry (AB, AC), and equivalence (BC, CB) trial types, appeared in the same block of 30 trials, as is shown in Table 1 (top panel). The testing blocks for the hybrid SIM–STC protocol included separate blocks for each symmetry (AB and then AC) and another block for the two (AB–AC) symmetry trial types, before the final equivalence (BC–CB) test block, as is also shown in Table 1 (top panel). No simple transitive testing is possible in the CAN training structure, since the BC–CB relations represent the combined test for symmetry and transitivity (Saunders & Green, 1999). In a within-participant comparison, all participants experienced both the standard SIM–SIM and hybrid SIM–STC protocols, although in different orders, during the same session.

Table 1 Sequence of training and testing blocks in each condition of Experiments 1 (top panel) and 2 (bottom panel)

| Condition	Block(s)	Relation	Feedback	Trials/Block
Experiment 1				
SIM–STC (hybrid)	Training	BA, CA	100, 50, 0	12 each
	Testing	AB, AC	0	9 each
		AB–AC	0	12
		BC–CB	0	20
SIM–SIM (standard)	Training	ED, FD	100, 50, 0	12 each
	Testing	DE, DF, EF, FE, ED, FD	0	30
Experiment 2				
SIM–STC (hybrid)	Training	BA, CA	100, 50, 0	12 each
	Testing	AB, AC	0	9 each
		AB–AC	0	12
		BC–CB	0	24
SIM–SIM (standard)	Training	ED, FD	100, 50, 0	12 each
	Testing	DE, DF, EF, FE, ED, FD	0	54

Results and discussion

A paired t test revealed significantly better performance [$t(15)$ = 3.41, $p < .05$, $\eta^2 = .44$, $d = 1.1$] for the overall percentage of accuracy on all training and testing blocks under each protocol for the hybrid SIM–STC ($M = 89.3$, $SD = 6.5$, $N = 16$) than for the standard SIM–SIM ($M = 81.4$, $SD = 7.8$, $N = 16$) protocol. Figure 2 presents the percentages correct on baseline relations across the different levels of feedback. Similar patterns of acquisition occurred under the two protocols, indicating no systematic order effect. Notably, four (25 %) of the participants (Kano, Ondo, Ogun, and Iwa) completed substantially more training blocks before achieving criterion under 100 % feedback in the standard SIM–SIM than in the hybrid SIM–STC protocol; only Milo completed substantially more blocks in the hybrid than in the standard protocol. Most participants (63 %) under the 50 % and 0 % feedback levels met criterion in one block, with the exception of Ebo with 50 % feedback (SIM–STC), Iro and Ekiti with 0 % feedback (SIM–STC), and Ado, Ijo, and Kano with 0 % feedback (SIM–SIM). All participants achieved criterion on the training blocks before testing began.

All participants required only one testing block for symmetry (separate and combined) and equivalence during STC testing in the hybrid protocol, except two who completed more testing blocks (combined AB–AC for Ado, and AC and combined AB–AC for Ekiti). In the SIM–SIM protocol, seven participants (44 %; Ebo, Ese, Iwa, Milo, Oro, Ado, and

Iro2) completed more than one block during equivalence testing. Figure 3 presents the percentages correct under the two protocols on all baseline, symmetry, and equivalence test trials, totaling 50 trials in the hybrid SIM–STC condition and 30 trials on the first test block in the standard SIM–SIM condition for each participant. The figure shows that all participants (100 %) achieved the 90 % accuracy criterion demonstrating equivalence under the hybrid SIM–STC protocol, but only 12 participants (75 %) did so under the standard SIM–SIM protocol. The four participants (Oro, Ebo, Iro2, and Ado) who failed to demonstrate equivalence in the latter protocol exhibited more errors on symmetry and equivalence trials, as is shown in Fig. 4. In contrast, under the SIM–STC protocol, the limited number of errors were on baseline (B1A1for all, and C1A1 additionally for Ado) trials (not shown in Fig. 4). Thus, although these findings demonstrate equivalence under the hybrid protocol, and despite good performance on the baseline trials in the test blocks, errors by these participants variously occurred on symmetry and equivalence trials in the standard SIM protocol, resulting in the observed shortcomings on equivalence formation. The pattern of errors in the present experiment is consistent with those reported previously by Arntzen and Nikolaisen (2011), in which participants who formed equivalence exhibited no errors on baseline and symmetry trials, and those who did not form equivalence either made errors on all trial types or on baseline, symmetry, or both trial types.

The results of the present experiment thus show that the hybrid of SIM training with STC testing engendered better overall accuracy than did the standard SIM–SIM protocol, and with a 25 % difference in equivalence formation between the two protocols, they indicate better equivalence outcomes for the hybrid protocol on the whole, implicating the order of testing as a significant variable in the observed stimulus equivalence outcomes. Because the standard SIM protocol presents all derived equivalence relations typically in one large testing block instead of testing serially from simpler to more difficult derived relations, as is customary with the standard STC protocol, participants have been less likely to demonstrate equivalence using it (Buffington et al., 1997). Although the equivalence yield reported in the present experiment for the standard SIM–SIM protocol is higher than that reported (at about 60 %) by Buffington et al. when no pretraining was provided, perhaps because of the difference in the numbers of comparisons used (two there vs. three here; see Imam, 2006), the serial, cumulative testing following each baseline training that is characteristic of the STC protocol may be responsible for its success in fostering the positive equivalence outcomes that have previously been reported and were observed in the hybrid SIM–STC protocol of the present experiment.
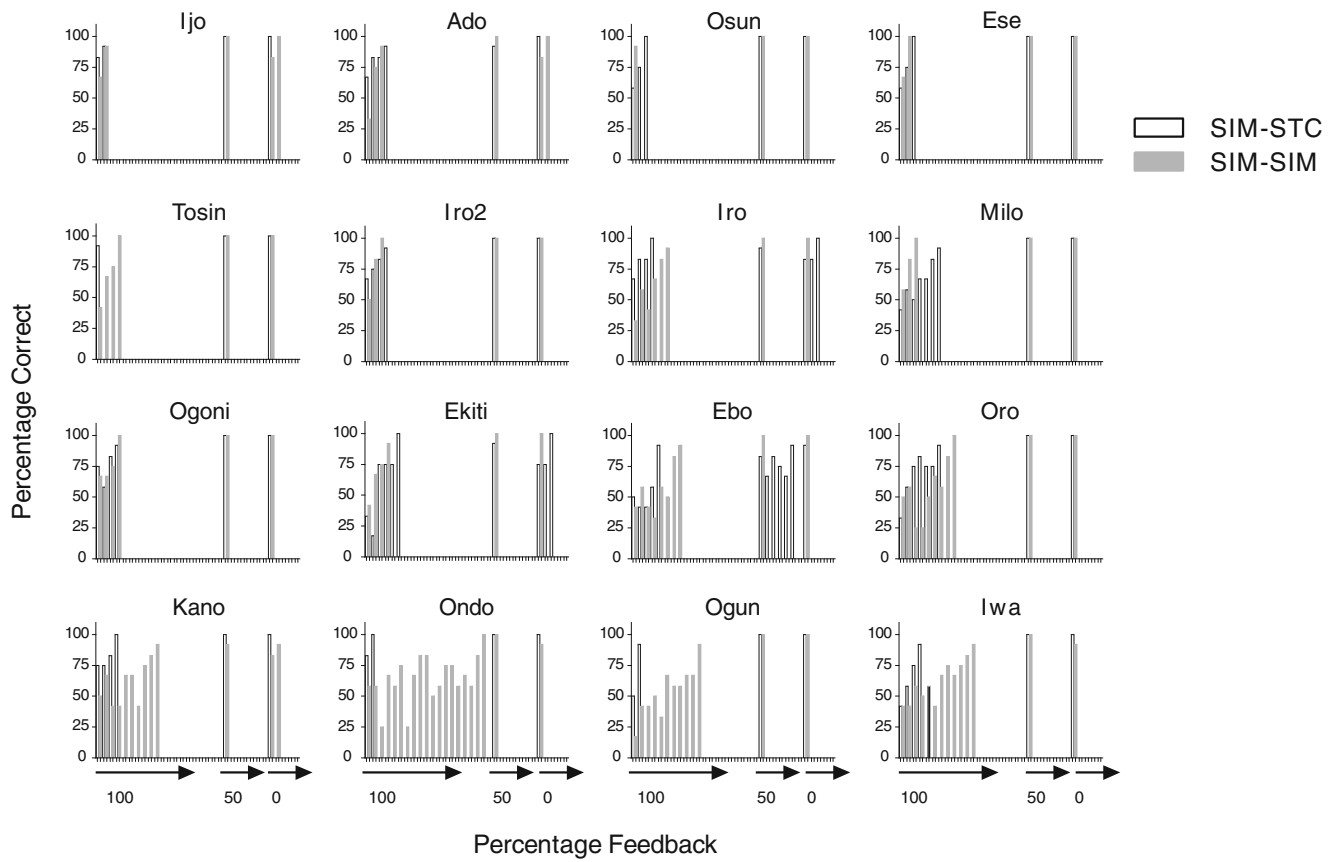
Fig. 2 Percentages correct as a function of training blocks under 100 %, 50 %, and 0 % feedback for each participant during training in Experiment 1. The left panels present data for participants who experienced the hybrid (SIM–STC) before the standard (SIM–SIM) protocols, whereas those on the right are for participants who experienced the standard before the hybrid
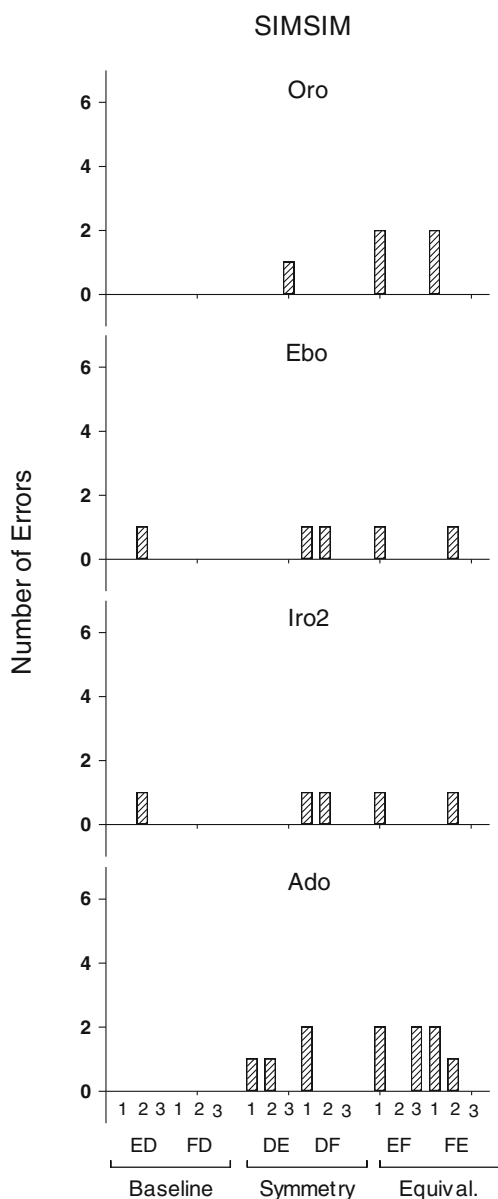
Experiment 2

In the previous experiment, two aspects of the stimuli used posed potential problems for the outcomes. First, the numbers of trials in the two protocols were not equal and could have contributed to the results. Second, the stimuli arranged under the two protocols were not counterbalanced; that is, all participants experienced the SIM–STC and SIM–SIM stimuli only in their respective conditions. In other words, the SIM–STC stimuli were never deployed for the SIM–SIM protocol, and vice versa. Without such counterbalancing, the respective outcomes engendered by the protocols could potentially have
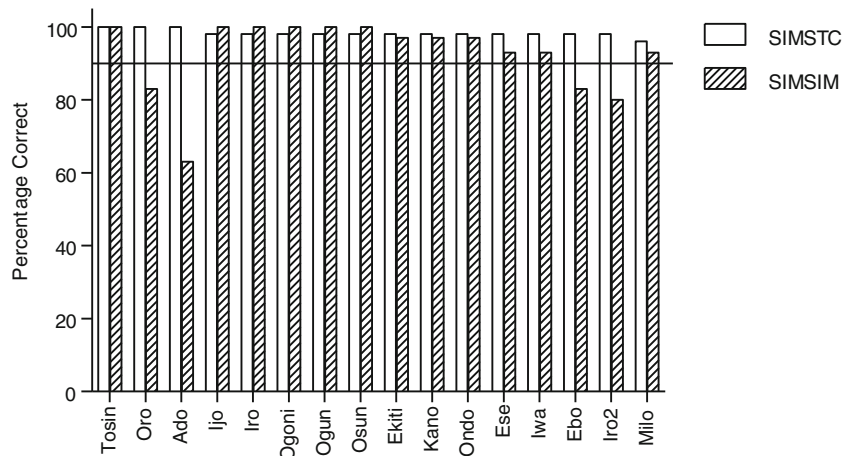
Fig. 3 Percentages correct on baseline, symmetry, and equivalence test trials combined for each participant in Experiment 1, under the hybrid SIM–STC and the standard SIM–SIM conditions

Fig. 4 Numbers of errors on each trial type during equivalence tests under the standard SIM–SIM for the four participants who did not meet criterion for equivalence formation in Experiment 1

been influenced by the particular stimuli used for each protocol. Previous research has implicated stimulus features as influencing equivalence outcomes (Fields, Arntzen, Nartey, & Eilifsen, 2012). In the present experiment, we sought to rectify these problems by (1) presenting equal numbers of trials under the two protocols and (2) randomizing allocation of the same stimulus sets that had been used in the previous experiment (see Fig. 1) into the two protocols implemented. In doing so, the unique characteristics of each stimulus and the particular configuration of stimulus combinations in the previous sets were randomly dispersed and reconfigured in no particular order in the present experiment.

Method

Participants

The participants consisted of 14 undergraduate students, six males and eight females, who were mostly freshman, in addition to two sophomores and three juniors. The participants ranged between 18 and 21 years of age and were enrolled at John Carroll University. Two other participants withdrew before completion of the experiment. All of the participants had no previous research experience, except two who reported nonoperant research experiences. They had not participated in the previous experiment.

Apparatus and stimuli

These were as described in Experiment 1. Figure 5 presents the stimulus sets for each protocol, randomized from those in the previous experiment.

Procedures

The procedures were similar to those of Experiment 1, with the exception that the same stimulus sets for the two conditions were selected randomly from those used in the previous experiment into the sets used for the hybrid and the standard conditions in the present experiment. As in the previous experiment, all of the participants experienced both protocols, but in different orders. The number of trials in the SIM test blocks was increased to 54 (from 30 trials in the previous experiment) to match those in the hybrid test blocks (see Table 1, bottom panel). Incorrect responses on five trials in the 12-trial training blocks automatically initiated retraining with 100 % feedback.

Results and discussion

Consistent with the previous experiment, a paired t test revealed significantly better overall accuracy [$t(13) = 2.42$, $p < .05$, $\eta^2 = .31$, $d = 0.5$] for the participants in the present experiment with the hybrid SIM–STC ($M = 77.5$, $SD = 9.6$, $N = 14$) than with the standard SIM–SIM ($M = 73.3$, $SD = 9.5$, $N = 14$) protocol. As in the previous experiment, we found no systematic order effect on acquisition. Three participants (21 %) completed substantially more training blocks in, respectively, the standard SIM–SIM protocol (Dido, Bo, and Kola) and the hybrid SIM–STC protocol (Biu, Elva, and Gora) than in the other protocol. Five of the participants (36 %; Ajide, Deba, Elva, Gora, and Biu) completed more blocks under the hybrid protocol with 100 % feedback, but only two participants (14 %; Abid and Oron) completed equal numbers of blocks (see Fig. 6). For the other feedback levels, only 14 % of the participants (Ikire
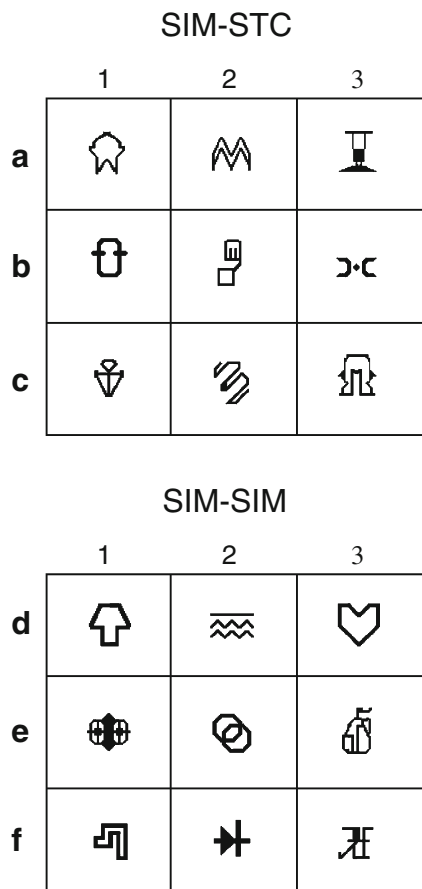
SIM-STC

SIM-SIM

Fig. 5 Stimuli used in the hybrid SIM–STC (top) and the standard SIM–SIM (bottom) protocols in Experiment 2. The letters designate sets of sample and comparison stimuli, and the numbers designate potential stimulus classes

and Biu with 50 % feedback, and Oron and Seyi with 0 % feedback) completed more than one block in the hybrid protocol, and 14 % (Abid and Kodi with 50 % feedback) completed more than one block in the standard protocol. All participants achieved criterion on the training blocks before testing began.

In addition, four participants (Abid, Biu, Bo, and Ikire) completed more than one block of AC symmetry testing. All of the participants completed two blocks of BC–CB equivalence testing in the SIM–STC protocol and two testing blocks in the SIM–SIM protocol. Figure 7 shows that the equivalence outcome was better under the hybrid SIM–STC than under the standard SIM–SIM protocol. Not factoring in Dido, who achieved 89 % accuracy on all trials, eight others (57 %) in the hybrid SIM–STC protocol achieved criterion performance demonstrating equivalence, as compared to 0 % in the SIM–SIM protocol. The 100 % difference in equivalence formation in the hybrid SIM–STC over the standard SIM–SIM protocol in the present experiment supports the 0 % equivalence yield under standard SIM protocols previously reported (Fields et al., 2012; Fields et al., 1997). The results vindicate those of Experiment 1, because any stimulus aspects, such as the

number of trials implemented or the particular stimuli in the stimulus sets used, that might have contributed to the stimulus equivalence outcomes in that experiment were controlled for by equalizing the number of test trials and randomizing stimulus allocation to the potential equivalence class sets in the present experiment. Nevertheless, this resulted in better hybrid SIM–STC protocol effects than were achieved in the standard SIM–SIM protocol.

Unlike in the previous experiment—in which a ceiling effect, particularly in the hybrid protocol, precluded it—these results provide ample opportunities to compare more directly failures to form equivalence under the two protocols. The error patterns, depicted in Fig. 8, are particularly useful in this regard. First, all successful equivalence outcomes (Ajide, Abid, Deba, Garo, Kodi, Kola, Ikire, and Sabo) under the SIM–STC protocol involved minimal or no errors on baseline (BA, CA) and symmetry (AB, AC) trials, a finding that is consistent with previous reports (Arntzen & Nikolaisen, 2011; Saunders, Drake, & Spradlin, 1999). Second, all failures to form equivalence under the two protocols involved more errors on both baseline (BA, CA) and symmetry (AB, AC) trials—as in the previous reports by Fields et al. (2000)—than did their successful counterparts, unlike in Experiment 1 (see Arntzen & Nikolaisen, 2011, for data similar to those of Exp. 1). Third, all failures to form equivalence under the SIM–STC protocol (Dido, Seyi, Bo, Elva, Biu, and Oron) involved errors across the three trial types, a finding not reported previously. Fourth, although this is sometimes the case with failures under the SIM–SIM protocol as well (Garo, Elva, Oron, Dido, Seyi, and Bo), some equivalence failures occurred under the SIM–STC protocol without errors on equivalence (BC, CB) trials (i.e., Ajide, Abid, Deba, Biu, Ikire, and Sabo), despite multiple errors on baseline and symmetry trials. Similar findings of equivalence without symmetry, baseline, or both trial types have been reported in the literature (Arntzen & Nikolaisen, 2011; Eilifsen & Arntzen, 2009; Saunders et al., 1999).

Completing equivalence trials successfully, while failing on symmetry trials, reflects the relative independence of stimulus equivalence properties that others have reported previously (Pilgrim & Galizio, 1996; Swisher, 2010). The differences in error patterns across the two protocols in the present experiment raise questions about the extent to which features of their tests are responsible for the various outcomes observed here, independent or not. Is it possible that the training, which was the same for both conditions in this and the previous experiment, was contributory to these outcomes? Arntzen and Nikolaisen (2011) employed a serial training procedure in a SIM protocol in which they trained each baseline relation in separate blocks before testing for any emergent relation with MTO and OTM training structures. A similar procedure was implemented by Saunders et al. (1999) in their first experiment. Both studies reported few
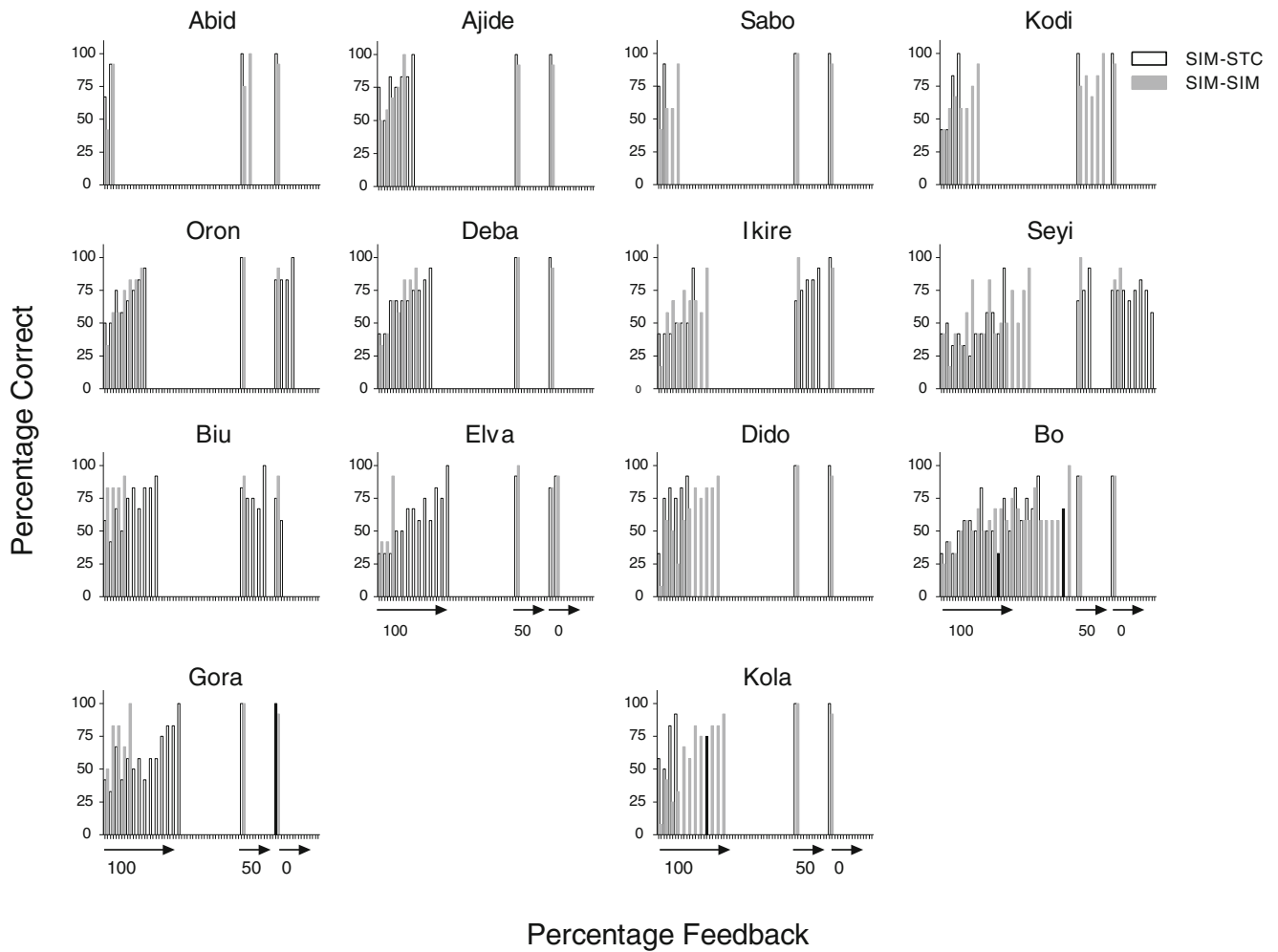
Fig. 6 Percentages correct as a function of training blocks under 100 %, 50 %, and 0 % feedback for each participant during training in Experiment 2. The left panels 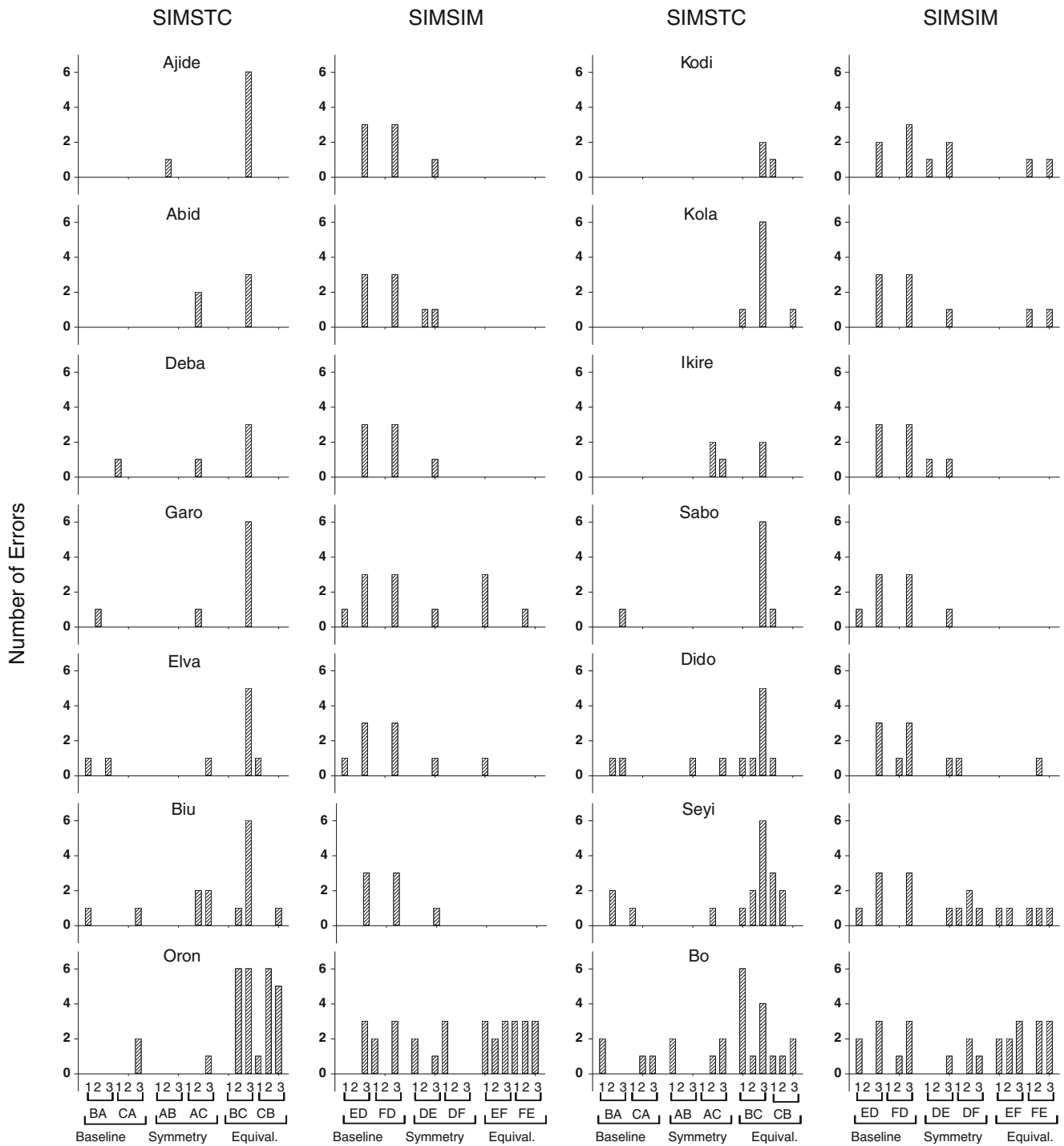present data for participants who experienced the hybrid (SIM–STC) before the standard (SIM–SIM) protocols, whereas those on the right are for participants who experienced the standard before the hybrid

participants who successfully completed equivalence trials while failing on symmetry and/or baseline trials. This aspect of their results suggests that the independence of the derived relations may be limited to serial training and/or combined testing in SIM protocols (Arntzen & Nikolaisen, 2011; Saunders et al., 1999; see also Eilifsen & Arntzen, 2009),
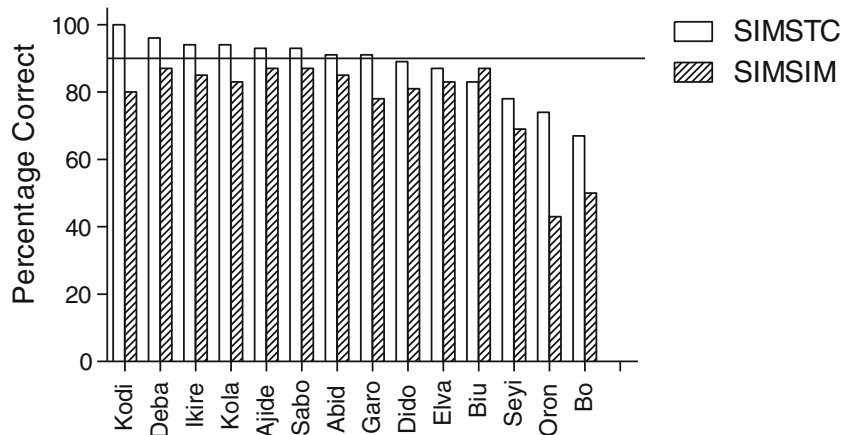
Fig. 7 Percentages correct on baseline, symmetry, and equivalence test trials combined for each participant in Experiment 2, under the hybrid SIM–STC and the standard SIM–SIM conditions

Fig. 8 Numbers of errors on each trial type during equivalence tests under the hybrid SIM–STC (left panel) and the standard SIM–SIM (right panel) conditions for all participants in Experiment 2. The left panels present data for participants who experienced the hybrid (SIM–STC) before the standard (SIM–SIM) protocol, whereas those on the right are for participants who experienced the standard before the hybrid

and to serial testing in the hybrid protocol of the present experiment. The differential impacts that this potential limitation of the disparate training regimes in each standard protocol would have for exploring whether or not equivalence success could be similarly possible in a standard STC protocol remains an empirical question, at this point.

General discussion

Previous studies of the simultaneous protocol have not addressed directly why this protocol tends to have a poorer equivalence-outcome record than the other two protocols, particularly the STC (Adams et al., 1993; Fields et al.,

1997). To fill this important gap in the literature, in the present study we focused on the test order of emergent relations, which occurred serially and cumulatively, as is typical of the STC protocol. This is in contrast to the collective, massed testing that is typical of SIM protocols, following regular SIM training. The results of the present study suggest that previous reports of poor yields and performance most likely were due to the massed nature of testing entailed by the standard SIM protocol. Because the standard SIM protocol presents all derived equivalence relations typically in a large block of testing, rather than testing serially from simpler to more difficult relations, as is customary with the STC protocol, participants have been less likely to demonstrate equivalence using it (Buffington et al., 1997; Fields et al., 1997).

The results of Experiment 2 are pertinent to the question of what about the standard SIM protocol predisposes it to failure more than the other protocols are. One clear pattern emerged in that experiment from the total failure to form equivalence under this protocol, similar to the 0 % yields previously reported by others (Fields et al., 2012; Fields et al., 1997): Participants' errors here were largely on baseline trials (and sometimes on symmetry trials), despite the participants having successfully achieved criterion performance during training. Conversely, participants' errors under the hybrid SIM–STC protocol tended to be mostly on emergent equivalence (and sometimes symmetry) trials with or without equivalence formation, despite their sharing a common training procedure with the standard SIM protocol. SIM testing appears to be the culprit in accounting for higher failure rates on requisite baseline and lower-order emergent (symmetrical) relations. The error patterns under the hybrid protocol would be expected by any theoretical standard, especially those that consider relational types to be hierarchically ordered (e.g., Adams et al., 1993; see Imam, 2001; and cf. Pilgrim & Galizio, 1996).

Another clear difference across the hybrid and standard protocols in Experiment 2 was that the SIM protocol engendered better performance on equivalence trials than on baseline and/or symmetry trials, a phenomenon that has been reported under similar conditions by others (Arntzen & Nikolaisen, 2011; Eilifsen & Arntzen, 2009; Saunders et al., 1999; Swisher, 2010). This pattern was remarkably absent throughout the hybrid SIM–STC protocol of the present study, however, and thus warrants further attention to similar data from the literature. Incidentally, the data that conform to the pattern of successful equivalence performance with failed symmetry and/or baseline performance were obtained under the standard SIM protocol. In Experiment 1 of Saunders et al.'s (1999) study, for example, three participants—Alice under MTO and Rita and Tina under OTM training structures—completed equivalence trials successfully while failing on symmetry trials. Eilifsen and Arntzen (2009) serially trained all baseline relations using an LS training structure

before the SIM tests; two participants (#3006 and #3019) succeeded on equivalence trials but failed on transitive trials, and another participant (#3010) succeeded on equivalence and transitive trials but failed on both baseline and symmetry trials (see their Fig. 2, p. 194). Also using an LS structure, Swisher (2010) deployed all baseline training together, as in the present study; a participant (S19) completed transitive trials successfully, while failing on symmetry and equivalence trials. Using training similar to that used by Eilifsen and Arntzen with SIM testing, Arntzen and Nikolaisen (2011) used MTO and OTM training structures with familiar and abstract stimuli; two participants (#7811 with abstract stimuli and #7823 with familiar stimuli) completed equivalence trials successfully while failing on symmetry trials (see their Table 5, p. 113).

These data, in conjunction with those of the present study, all seem to provide additional empirical support for what Pilgrim and Galizio (1996) referred to as "differential emergence of equivalence-defining properties," in which one derived relation, such as symmetry, does not serve as a requirement for another, like transitivity or equivalence, in a way that denotes "an integrated unit" of equivalence properties (p. 183). It appears that a by-product of the SIM protocol is a scrambled rather than an ordered "hierarchy" of the properties of equivalence. What the STC format of testing appears to accomplish is to impose a systematic order on what comes out of the equivalence-class urn (to adopt the "bag" analogy; Sidman, 2000), culminating in a differentiated ordering of classes in which all emergent relations would prevail, if any did. To the extent that testing arrangements may determine the outcome of equivalence formation (Sidman, 1994; Sidman et al., 1985), in the absence of an ordered presentation of test trials and with no feedback on their correctness, as would be the case in a SIM-protocol test, the participant makes choices among the comparisons without the kinds of differentiations afforded by the STC testing regime. This results in a scrambled ordering of classes in which any of the emergent relations may manifest, if any at all do.

Although the present study has focused mainly on testing procedures, whether or not certain aspects of training in the STC protocol also played important roles in producing better success than in the SIM protocol remains to be empirically demonstrated. In contrast to the SIM training used in the present experiments, what would happen if the respective standard training regimes of each protocol were implemented? Would this reveal a limited training effect? Would the pattern of errors be confirmatory of those obtained when common SIM training was used with different testing regimes in the two protocols of the present study?

Demonstrating experimental control is the hallmark of a behavioral analysis. These controls often are accomplished in different ways, such as through contingency, stimulus control, and instructional control, to name a few options. Stimulus equivalence research has benefited from a variety of them as

we seek to gain a fuller understanding of the phenomenon. In the present study, we attempted to explore the difficulties that many have reported under the simultaneous protocol, by relying largely on contingency manipulations with minimal instructions. The subtlety of the manipulation of the protocols within and across sessions appears to have had an impact on the equivalence outcomes, presumably by suggesting to participants that something was different. Undoubtedly, such difference(s) could be transmitted directly to participants via instructions. Nevertheless, to the extent that a contingency approach was adopted here, it remains to be seen what instructional control could produce the same effects obtained in the present study, or similar ones. We do not yet fully understand what aspects of the differences in STC and SIM protocols engender differential equivalence outcomes. Future research should explore the debate on the differential effectiveness of OTM versus MTO (Arntzen & Holth, 2004; Hove, 2003). For example, it remains to be seen what role using training structures other than the MTO structure used in the present study might reveal about differential equivalence outcomes when combined with these protocols.

References

Adams, B. J., Fields, L., & Verhave, T. (1993). Effects of the test order on the establishment and expansion of equivalence classes. *Psychological Record, 43,* 133–152.

Arntzen, E., & Holth, P. (2004). Equivalence outcomes in single subjects as a function of training structure. *Psychological Record, 50,* 603–628.

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.

Buffington, D. M., Fields, L., & Adams, B. J. (1997). Enhancing equivalence class formation by pretraining of other equivalence classes. *Psychological Record, 47,* 69–96.

Dube, W., & Hiris, J. (1997). *Matching to Sample Program (Version 11.08a67) [Computer software]*. Waltham, MA: E. K. Shriver Center for Mental Retardation.

Eilifsen, C., & Arntzen, E. (2009). On the role of trial types in tests for stimulus equivalence. *European Journal of Behavior Analysis, 10,* 187–202.

Fields, L., Arntzen, E., Nartey, R. K., & 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.

Fields, L., Landon-Jimenez, D. V., Buffington, D. M., & Adams, B. J. (1995). Maintained nodal-distance effects in equivalence classes. *Journal of the Experimental Analysis of Behavior, 64,* 129–145.

Fields, L., Reeve, K., Rosen, D., Varelas, A., Adams, B., Belanich, J., & Hobbie, S. (1997). Using the simultaneous protocol to study equivalence class formation: The facilitating effects of nodal number and size of previously established equivalence classes. *Journal of the Experimental Analysis of Behavior, 67,* 367–389.

Fields, L., Varelas, A., Reeve, K., Belanich, J., Wadhwa, P., Derosse, P., & Rosen, D. (2000). Effects of prior conditional discrimination training, symmetry, transitivity, and equivalence testing on the emergence of new equivalence classes. *Psychological Record, 50,* 443–466.

Hove, O. (2003). Differential probability of equivalence class formation following a one-to-many versus many-to-one training structure. *Psychological Record, 53,* 617–634.

Imam, A. A. (2001). Speed contingencies, number of stimulus presentations, and the nodality effect in equivalence class formation. *Journal of the Experimental Analysis of Behavior, 76,* 265–288.

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.

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). New York, NY: Elsevier Science.

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

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

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.

Sidman, M., Kirk, B., & Willson-Morris, W. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior, 43,* 21–42.

Sidman, M., & Tailby, W. (1982). Conditional discriminations vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22.

Swisher, M. J. (2010). *Determining the relation between the moment of acquisition of baseline conditional discriminations and the emergence of equivalence relations* (Unpublished master's thesis, University of North Texas). Retrieved from http://digital.library.unt.edu/ark:/67531/metadc30515/m2/1/high_res_d/thesis.pdf