The Psychological Record, 2000, 50, 305-337

REACTION TIMES AND THE EMERGENCE OF CLASS CONSISTENT RESPONDING: A CASE FOR PRECURRENT RESPONDING?

PER HOLTH and ERIK ARNTZEN
University of Oslo, Norway

Three experiments explored reaction times and the emergence of consistent responding during tests for equivalence following prerequisite conditional discrimination training for the establishment of three 3-member classes. Using a simultaneous testing protocol following linear series training, Experiment 1 investigated differential reaction times to baseline and to test trials, and the consistency of emergent test performances. Only 1 of 10 subjects responded in accord with stimulus equivalence, and 2 subjects responded in accord with other consistent patterns. Reaction times to test trials were longer than to baseline trials, and there was a tendency for lower reaction times to be associated with consistent responding during testing, whether or not the consistent responding was in accord with equivalence. Experiment 2 investigated reaction times and the emergence of consistent responding during training as a function of repeated training and testing. One subject from the previous experiment participated in eight replications with the stimulus materials from Experiment 1 and two replications with a second set of materials. Neither of these produced stimulus equivalence, even after baseline and symmetry test performances were nearly perfect. However, with both sets of materials, reaction times were gradually reduced as responding became consistent with a different pattern during "equivalence" testing. To examine whether the longer reaction times initially during testing are directly relevant to the emergence of consistent responding and, under favorable conditions, stimulus equivalence, the opportunity to respond to comparison stimuli was restricted to 2 s in Experiment 3. The results showed that with the reaction time constriction, none of the subjects responded in accord with equivalence under otherwise favorable conditions, that is, following one-to-many training. A problem-solving interpretation of equivalence formation is suggested.

Thanks to Barry Lowenkron for comments on parts of an earlier version of the manuscript. Correspondence and reprints requests may be sent to Per Holth, Department of Psychology, University of Oslo, PO Box 1094 Blindern, 0317 Oslo, Norway. (E-mail: Per.Holth@psykologi.uio.no).

In current usage (after Sidman & Tailby, 1982), a minimal arrangement necessary for the testing of stimulus equivalence may include the following conditional discrimination training: Of two simultaneously present comparison stimuli, B1 and B2, the selection of B1 is reinforced when A1 is the sample, while the selection of B2 is reinforced when A2 is the sample. Next, when either B1 or B2 is presented as a sample, C1 or C2, respectively, serves as the correct comparison stimulus. Stimuli are considered as members of an equivalence class when their interrelations in a matching-to-sample task have the properties of reflexivity (e.g., if A1-B1 and A2-B2, then A1-A1, A2-A2, etc.), symmetry (e.g., if A1-B1 and A2-B2, then B1-A1 and B2-A2), and transitivity (e.g., if A1-B1 and B1-C1, then A1-C1). Following A-B and B-C training, a C-A test is a combined symmetry and transitivity test, and has been called an abbreviated equivalence test (Sidman, 1994).

According to Sidman (1997), the generality of stimulus equivalence has been shown in many laboratories, with many different kinds of stimuli, with varying numbers and sizes of classes, and with various teaching and testing procedures. Thus, Sidman (1990) suggested that stimulus equivalence may be a fundamental, "unanalyzable" stimulus function. In fact, the high reliability of class formation that results from some commonly used procedures with human subjects has been considered as "a major impediment to the identification of historical variables that influence the emergence of new equivalence classes by typically functioning adults" (Fields, Reeve, Rosen, Varelas, Adams, Belanich, & Hobbie, 1997, p. 368). Similarly, Saunders, Saunders, Williams, and Spradlin (1993) suggested that stimulus equivalence may be underanalyzed rather than "unanalyzable," and that it may remain so until subjects do not exhibit the predicted test performances until they are exposed to specifiable training histories. A starting point, then, is to identify exceptions to the pattern of test performance that defines equivalence. Another potentially fruitful path to the investigation of emergent performances as a function of training histories lies in the analysis of reaction times or speed, because this measure can be sensitive to different variables even when accuracy of responding is not (e.g., Spencer & Chase, 1996).

Variables Relevant to the Probability of Class Formation

Even though there are exceptions to the pattern of test performance that defines equivalence, Sidman (1993) argued that the behavioral tests for equivalence have proven so consistent that the most constructive response to apparent irregularities is a search for artifacts. A number of artifactual results have been summarized by Stikeleather and Sidman (1990). They include (a) unintended identity relations, in which experimenter-planned equivalence relations can be overridden by physical resemblances between stimuli, (b) a failure to establish the necessary prerequisites, which may result from insufficiently rigorous criteria for the acquisition of the basic conditional discriminations, (c) S-minus control, in which subjects reject incorrect comparisons rather than select the correct one and, hence, may show zero-scores on an equivalence test although a consistent controlling relation exists between the sample and a negative comparison, and (d) contextual control that obscures potential equivalence relations when stimuli are members of different equivalence classes or are also related to each other in various ways.

There are other irregularities to the equivalence pattern as well. The likelihood of class formation can be differentially affected by training design or structure (Arntzen & Holth, 1997; Barnes, 1994; Saunders, et al., 1993; Saunders, Wachter, & Spradlin, 1988; Spradlin & Saunders, 1986). For instance, training according to either of three different class structures can give rise to three-member classes: (a) *linear series* (AB and BC), (b) *many-to-one* (comparison-as-node; AB and CB), or (c) *one-to-many* (sample-as-node; AB and AC). Indeed, following the prototypical linear series, AB and BC training, with Greek letters as stimuli, the probability of experimenter-defined equivalence-class formation may not exceed chance (Arntzen & Holth, 1997).

The probability of the relevant emergent relations can also be a function of training and test protocol, that is, the order in which baseline training and testing for emergent relations are arranged. For instance, Adams, Fields, and Verhave (1993) distinguished between simple-to-complex and complex-to-simple protocols. In the *simple-to-complex* protocol, after training the conditional relation AB, a BA symmetry test is administered, then a new conditional relation, say BC, is directly trained, and a CB symmetry test is conducted before an AC transitivity test and, finally, a CA equivalence (i.e., combined symmetry and transitivity) test. In contrast, the *complex-to-simple* protocol establishes all the prerequisite conditional relations before testing and, then, tests combined symmetry and transitivity (equivalence) before transitivity and, finally, symmetry are tested separately. In a third type of training and test sequence, called the *simultaneous* protocol (e.g., Fields, Landon-Jimenez, Buffington, & Adams, 1995), all baseline relations are established before the administration of a mixed test block, consisting of a random sequence of symmetry, transitivity, and equivalence probes. Summarizing the results of several published reports of studies that used different training protocols, Fields et al. (1997) found that the simple-to-complex protocol was the most effective, and the simultaneous protocol was the least effective procedure in terms of percentage of subjects that demonstrate class consistent performances. For instance, Buffington, Fields, and Adams (1997) assessed the likelihood of class formation in college students under the simultaneous protocol following linear series training aimed at establishing two 3-member classes of nonsense words. They found that only 4 of 12 subjects responded in accord with the experimenter-defined classes during the first test block, while 2 more subjects responded in accord with this pattern after test block repetitions. In contrast, experiments using the simple-to-complex testing protocol typically have reported the emergence of equivalence classes in 93-100% of the subjects (Fields et al., 1997).

Furthermore, the probability of equivalence class formation can be a function of stimulus familiarity (Holth & Arntzen, 1998a) and of other features of experimental or preexperimental histories. For instance, both the number of nodes and the number of members of previously established stimulus classes can facilitate the subsequent establishment of new performances that are in accord with equivalence (Fields et al., 1997). The reverse effect has been reported following histories that involve unidirectional rather than bidirectional relations, as with letters in alphabetical order (Holth & Arntzen, 1998b). Moreover, Wulfert, Dougher, and Greenway (1991) found that whereas prior training to attend to the relations among pictorial stimuli enhanced the likelihood of subsequent equivalence class formation, prior training to attend to pictorial stimuli as elements of a compound reduced the probability of forming new classes.

Variables Relevant to Reaction Times

Spencer and Chase (1996) suggested another research path aimed at the more precise prediction and control of emergent relations and, thus, to the study of differential effects of training histories, namely by measuring reaction times or response speed. Several studies have shown that reaction times may be sensitive to a number of variables even when accuracy of responding is similar across trial types. Typically, reaction times to comparison stimuli initially during testing are markedly longer both than during the last training trials and during later test trials (Arntzen & Holth, 1997; Holth & Arntzen, 1998a). A study by Wulfert and Hayes (1988) showed longer reaction times to combined tests than to baseline and to symmetry test trials. Bentall, Dickins, and Fox (1993) showed that reaction times were longer on matching tasks that involved stimuli that had not been directly related during training than on tasks that involved previously directly related stimuli. Spencer and Chase (1996) extended these results and found that responding was faster on symmetry test trials than on both transitivity and combined test trials, but faster yet on baseline trials.

No studies seem to have been directly concerned with whether differential reaction times to different trial types are related to whether or not the experimenter-predicted performances actually emerge. Hence, the purpose of the present study was to investigate relations between reaction times and the probability of class consistent responding during testing: First, do differential reaction times to different trial types correlate with the actual emergence of class consistent responding? Second, will the probability of class-consistent responding be affected by a reaction time constraint initially during testing? One possibility is that the phenomenon of longer reaction times initially during testing is a relatively unimportant side effect of some novel stimulus constellations. Another possibility is that something highly relevant to patterns of responding during the test might occur during those trials with longer reaction times.

Experiment 1

In those previous studies that have shown low probabilities of equivalence class formation, the emergence of stimulus equivalence relations was measured during separate test blocks in extinction. With such a test-block procedure, one cannot decide whether stimulus equivalence fails to emerge from the established conditional discriminations, or whether those trained conditional discriminations, from which equivalence is expected to emerge, themselves break down. Therefore, the main purpose of Experiment 1 was to investigate the effect on the probability of equivalence class formation and on the stability of baseline performances when test probes are interpolated between intermittently reinforced training trials. This test probe procedure allows for the continuous monitoring of baseline performance during testing. Furthermore, we were interested in learning whether the previous findings of longer reaction times to test trials than to baseline trials, and the different patterns of consistent responding would be replicated in the test-probe procedure.

Method

Subjects

Ten staff members from a residential and treatment center for autistic youths served as subjects. None of the subjects had previously participated in research on stimulus equivalence.

Apparatus

A personal computer controlled stimulus presentation and data collection. The stimulus materials were Greek letters that were displayed on the monitor. The stimuli are shown in Table 1. A square (7x7 cm) on the left side of the monitor served as a sample stimulus key. Six squares (4x4 cm), arranged in two columns and three rows on the right side of the screen, served as comparison stimulus keys. A transparent touch screen mounted in front of the monitor measured the locations of subjects selections. A cassette player controlled by the computer played music immediately following correct responses.

Table 1

Stimulus Materials

	1	2	3
A	δ	φ	λ
B	γ	μ	ξ
C	ψ	ζ	θ

Procedure

Instructions. When asked to participate in the experiment, subjects were told that the experiment was concerned with learning processes and involved tasks presented on a computer with a touch screen. The subjects were also told that the experiment would last for approximately 40 min, depending on how rapidly and correctly they responded. Detailed instructions given when the subjects were seated in front of the monitor are shown in Appendix A.

Training and testing. Each trial started with the presentation of a sample stimulus on the left side of the screen. Touching this stimulus was followed by the presentation of comparison stimuli in the keys on the right side of the monitor. The sample remained until a comparison stimulus was touched. To minimize the number of errors occurring initially during training, the conditional discrimination tasks were introduced step by step as follows:

Successively over the first three trials, each of the A stimuli were presented as a sample on the left side of the screen. Touching the sample produced the presentation of the correct comparison stimulus on either of the six keys on the right side. Touching the comparison stimulus produced music from the cassette player for 2.0 s. During the next three trials, a touch on the sample produced both the correct and one of the two incorrect comparisons. An incorrect selection was followed by the blanking of the screen for 5 s before the procedure resumed from the start. After three successive correct comparison selections, the next three trials required the subject to select the correct comparison when presented simultaneously with the second incorrect one. During the last three introductory trials, touching the sample produced the presentation of the correct as well as both of the incorrect comparisons. From trial to trial, the three comparison stimuli appeared in randomly selected positions, except that there was never more than one comparison in each row of keys on the screen. Following this introduction of the comparison stimuli, AB training continued until the subject completed 21 successive correct trials. BC training was accomplished using the same procedure. Next, AB and BC trials were quasi-randomly intermixed, and training with them continued to a criterion of 24 consecutive correct trials. Finally, 20 symmetry test probes (BA and CB) and 20 equivalence test probes (CA) were quasi-randomly mixed with 38 training trials. During the test phase, differential consequences for correct and incorrect training trials were arranged according to a variable ratio 3 schedule. A correct selection was then followed by music as described above, whereas an incorrect selection led to the immediate blanking of the screen for 5 s and to the repetition of that particular trial until a correct comparison selection occurred.

Three sources of data were measured: Correct and incorrect selections, the reaction times of these selections, and the number of trials needed to reach criterion. The proportion of responding in accord with experimenter-defined class consistent responding required in order to consider the tests as positive was defined as minimum 9/10.[1]

The consistency of responding to equivalence test trials was evaluated throughout the test phase. Similar to a 'moving average,' a consistency index was calculated for each block of six successive

[1]This criterion may appear to be unreasonably strict since, given completely random selections across trials, the probability of minimum 9/10 is below 0.0005, and even for minimum 7/10, the probability is well below 0.05. However, truly random selection cannot be presumed. Consistent selection patterns can emerge even when they are not in accord with experimenter-defined classes. Furthermore, some subjects come to select different comparisons for each sample. On a three-choice task as in the present experiment, such selections would render a probability of 1/6 of hitting upon the experimenter-defined pattern. (Holth & Arntzen, 1998a)

equivalence test trials: (e.g., Trials 1-6, Trials 2-7, Trials 3-8 up to Trials 15-20.) Each such six-trials sequence included at least one trial with the sample from each class (i.e., C1, C2, and C3) and, therefore, a maximum of three selections could conflict with other selections within a six-trials sequence. A consistency index of 1.0 indicated that no comparison selections conflicted with another selection within that particular sequence of six trials. Each time a sample occasioned a comparison selection not in accord with other choices within that six-trial block, the consistency index would drop by 0.33. Hence, an index of 0.67 conveys that one comparison selection was made in conflict with another selection, an index of 0.33 shows that two selections diverged from other selections, and an index of 0.0 implies that the maximum of three comparison selections were in conflict with other selections within that particular sequence of six equivalence test trials.

Results

The introduction of test probes was accompanied by some breakdown of consistency of selections during training trials in most of the subjects, the number of errors ranging from 1 to 11 of 38 baseline AB and BC trials (see Table 2). Five subjects maintained baseline performances making four or fewer errors on the 38 baseline trials. Of these subjects, 3 responded in accord with BA and CB symmetry, but only 1 of them (#69) also responded in accord with CA equivalence. None of the subjects who did not respond in accord with symmetry responded in accord with "equivalence" on the combined test probes.

Table 2

Number of Trials and Errors During Training and Proportion of Correct Responses to Baseline Trials, Symmetry Test Probes, and Equivalence Test Probes During the Test Phase for all Subjects in Experiment 1

Training Phase			Test Phase							
			Baseline training trials				Symmetry probes		Equivalence probes	
Subject	Trials	Errors	1-10	11-20	21-30	31-38	1-10	11-20	1-10	11-20
69	130	3	**10/10**	8/10	**10/10**	**8/8**	**10/10**	**10/10**	7/10	**9/10**
66	129	3	**10/10**	**10/10**	**10/10**	**7/8**	**10/10**	**10/10**	7/10	8/10
61	96	2	**9/10**	8/10	8/10	6/8	**9/10**	**9/10**	7/10	7/10
68	146	8	**10/10**	8/10	8/10	**8/8**	8/10	**9/10**	2/10	5/10
67	222	18	**9/10**	8/10	7/10	**7/8**	8/10	8/10	2/10	4/10
70	108	3	**9/10**	**9/10**	**9/10**	**8/8**	8/10	7/10	4/10	8/10
65	156	12	8/10	**9/10**	7/10	**7/8**	6/10	8/10	3/10	1/10
63	96	1	**9/10**	6/10	8/10	4/8	7/10	7/10	1/10	3/10
62	348	37	**10/10**	8/10	**9/10**	**8/8**	6/10	7/10	2/10	2/10
64	192	22	**10/10**	**9/10**	6/10	**8/8**	4/10	7/10	3/10	6/10

Note. Subjects are arranged according to symmetry test scores. The results from baseline trials are split in four for the monitoring of changes in the persistence of the trained relations. Similarly, the results from the symmetry test probes and from the equivalence test probes are split half for the easy detection of delayed emergent relations.

Among the remaining 5 subjects, with five or more errors of the 38 baseline trials, only 1 subject (#61) responded in accord with symmetry, and none responded in accord with equivalence.

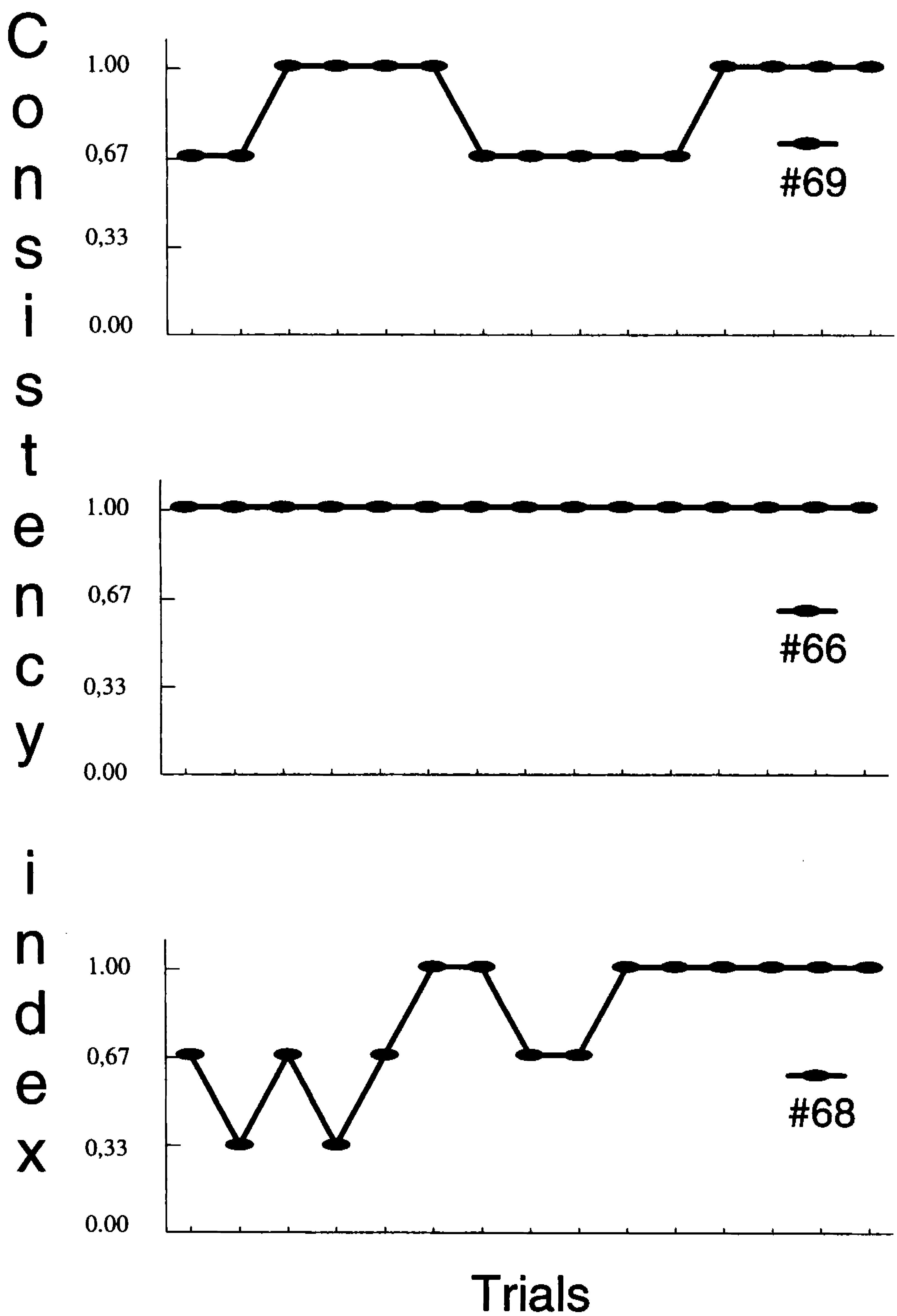

Figure 1. Successive consistency indices for the three subjects in Experiment 1 who responded consistently to equivalence probes towards the end of testing. Only Subject #69 responded in accord with equivalence.

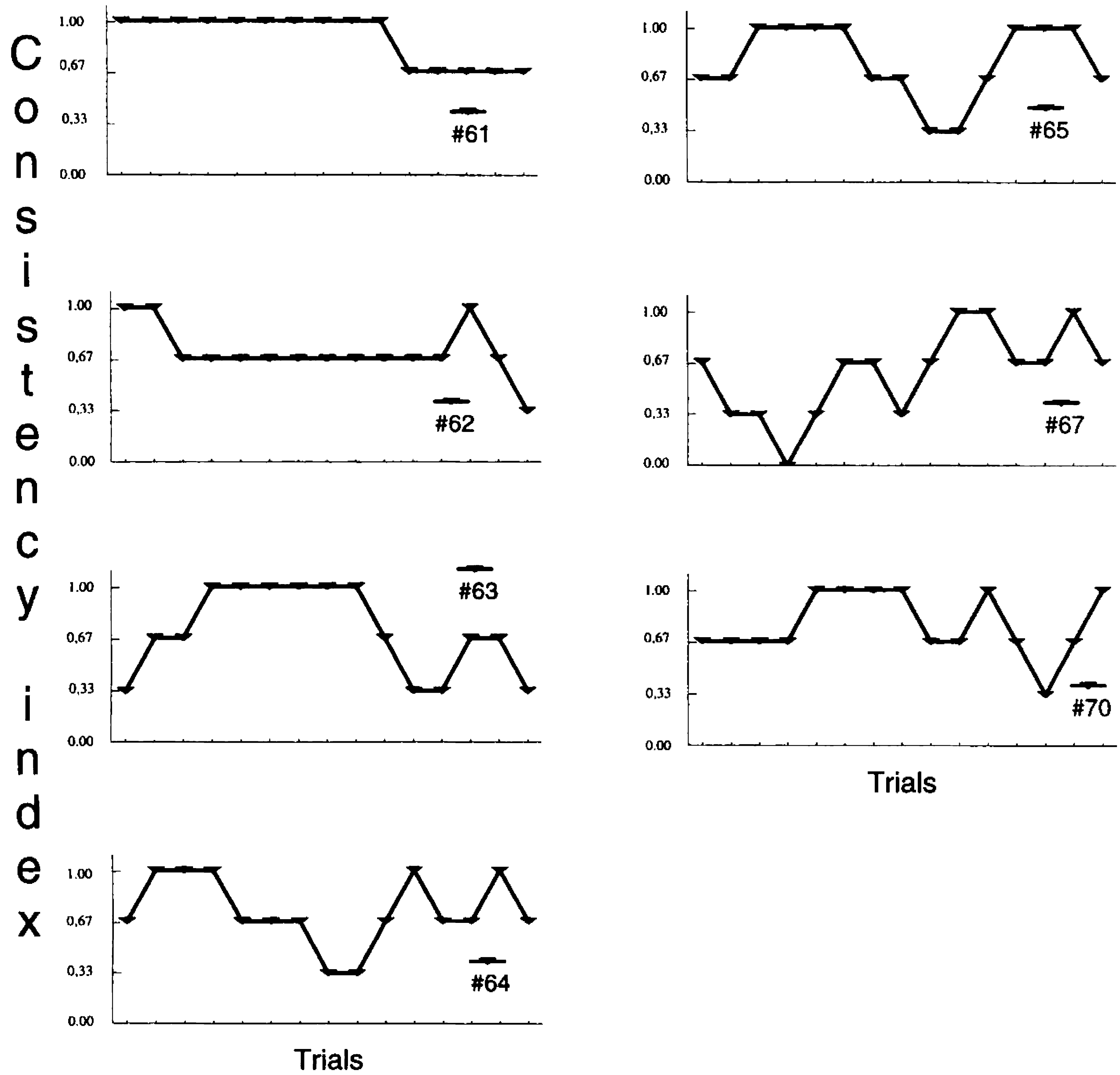

Figure 2. Successive consistency indices for the subjects in Experiment 1 who did not respond consistently to equivalence probes.

Of the 9 subjects who did not respond in accord with equivalence, 2 eventually responded in accord with other patterns. Thus, as shown in Figure 1, in addition to Subject #69 who, eventually, responded in accord with equivalence, Subject #66 responded consistently throughout equivalence testing, and #68 did so after some initial inconsistency. Subject #66 selected comparison A1 in the presence of C1, but A3 in the presence of both C2 and C3 throughout testing. Subject #68, after some delay, showed a consistent pattern of selecting comparison A2 in the presence of C1, A1 in the presence of C2, and A3 in the presence of C3. Figure 2 shows that the remaining 7 subjects responded inconsistently towards the end of testing. Yet, totally, 5 subjects who did not respond in accord with equivalence did respond consistently in accord with at least one of the three experimenter-defined sample-comparison pairs during "equivalence" test probes.

Individual cumulative records of reaction times to comparison stimuli throughout the test probe phase of the experiment are plotted in Figure 3. Of the 10 subjects, 9 had longer reaction times to both symmetry probes

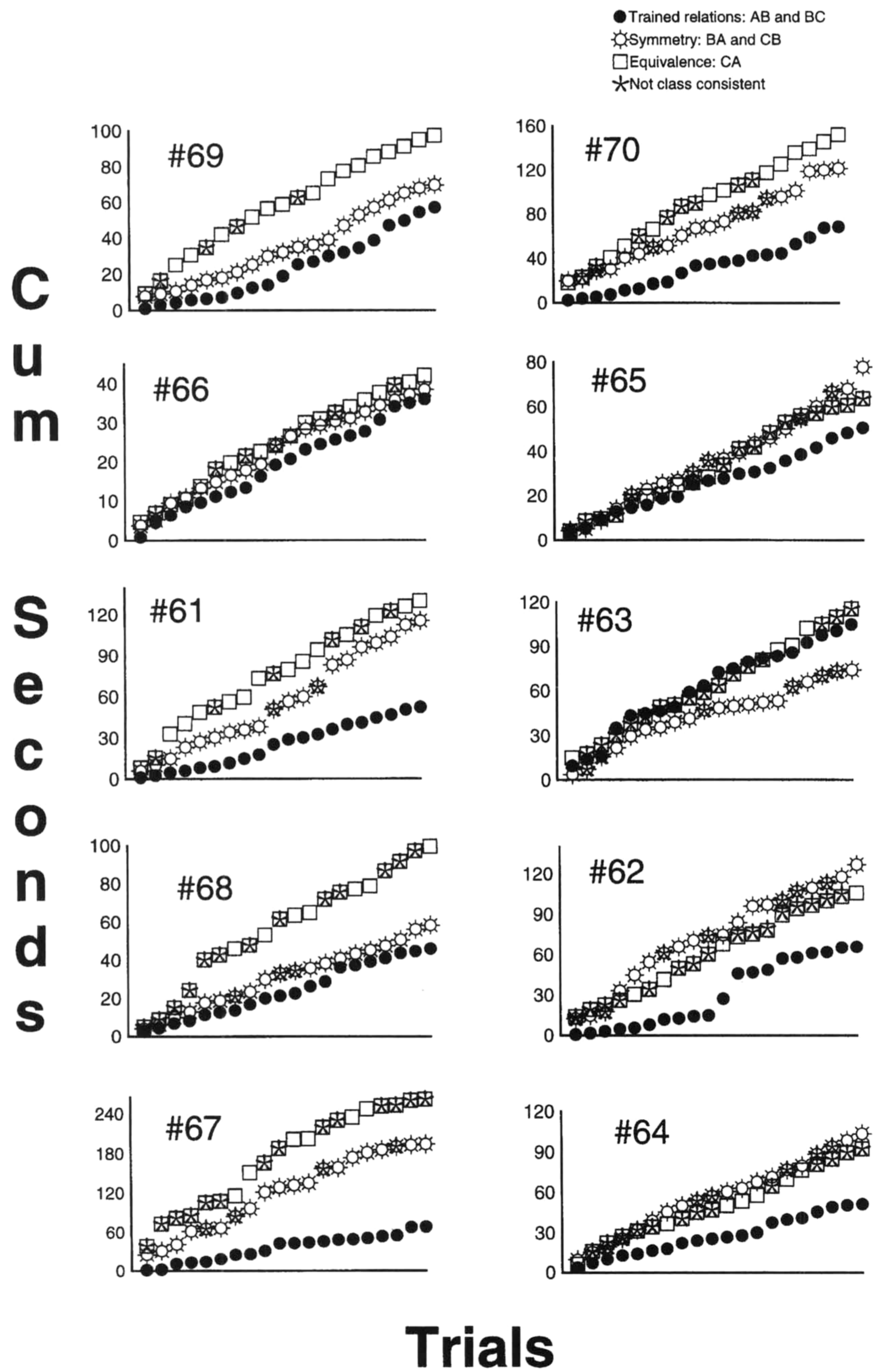

Figure 3. Individual cumulative reaction times to comparison stimuli during equivalence test probes, symmetry test probes, and the first 20 training trials in Experiment 1.

and to equivalence test probes than to baseline trials. Of these subjects, 6 also showed longer reaction times to equivalence test probes than to symmetry test probes. These 6 subjects were the ones with the highest symmetry test scores. Furthermore, the 4 subjects whose responding actually met the symmetry criterion were also those with the highest consistency indices for responding to equivalence test probes.

There were no conspicuous differences in reaction times within or across subjects for responding in accord with experimenter-defined classes versus for responding in discordance with those classes, as can be derived from Figure 3 (each test selection that was not consistent with experimenter-defined classes is marked with an asterisk in the figure). However, 3 of the 4 subjects with the shortest overall reaction times were those 3 subjects (#69, #66, and #68) who ended with a consistent pattern of responding during equivalence test probes (regardless of whether or not that pattern was the experimenter-defined one), as well as with responding in accord with symmetry. The 4th subject with relatively short overall reaction times (#65) responded inconsistently to all trial types throughout most of the test probe phase.

Discussion

The present experiment, using a type of simultaneous testing protocol in which test probes were interpolated between training trials, replicated the results of previous experiments that have used separate test blocks, both with respect to (a) a low probability of an equivalence outcome, and (b) the demonstration of longer reaction times to comparison stimuli during test trials than during baseline trials.

The likelihood of class formation reported here was similar to that reported by Arntzen and Holth (1997), when a separate equivalence test block was presented immediately following linear series training with Greek letters as stimuli. Relatively low yields were also obtained by Buffington et al. (1997) using a simultaneous testing protocol as in the present experiment. A common observation related to the simultaneous testing protocol is the disruption of baseline performances associated with the introduction of test probes (e.g., Buffington et al., 1997; Fields et al., 1995). Although the results of the present experiment indicate some correspondence between maintenance of baseline performance and class consistent performances, the question remains to what extent perfectly maintained baseline performances will be accompanied by perfectly class consistent responding during different test trials.

The yields in the present experiment were even lower than those reported by Buffington et al. (1997). Three obvious differences between the Buffington et al. study and the present one seem relevant in order to explain these differential outcomes. First, the study by Buffington et al. was based upon two-choice tasks, whereas the present experiment used three-choice tasks. It is possible that two-choice tasks will result in a higher probability of class consistent responding partly because it is simpler. There is also obviously the increased probability of false positives (cf. Sidman, 1987). Second, Buffington et al. used a start-up training procedure in which semantically related words were used as samples and comparisons along with instructions that explicitly told the subject to "discover which words go together." (cf. Fields, Adams, Verhave, & Newman, 1990). As pointed out by Sidman (1994), "Instructions to the subject may establish a context that brings into play historical contingencies that interact with or completely override current experimental contingencies" (p. 510). Third, the easily pronounceable nonsense syllables used by Buffington et al. versus the Greek letters used in the present experiment may have contributed to the differential yields (cf., Mandell & Sheen, 1994).

Of the 10 subjects, 9 had longer reaction times to both symmetry test probes and to equivalence test probes than to baseline trials. Of these subjects, 6 also showed longer reaction times to equivalence test probes than to symmetry test probes. Only 1 of these subjects responded in accord with equivalence on the combined test, and 4 of them responded in accord with symmetry. Thus, the present results replicated and extended those of Spencer and Chase (1996) by demonstrating differential reaction times to baseline, symmetry, and equivalence trials even when the equivalence relation did not emerge. However, these results diverge from those of Wulfert and Hayes (1988) and from those of Bentall et al. (1993) that did not show differential reaction times to baseline versus symmetry test trials. Spencer and Chase (1996) speculated that differences in reaction times to baseline and symmetry responding in these studies were "washed out" by the use of a 2-s delay between the termination of the sample and the presentation of comparisons. However, their own study also differed from those of Bentall et al. (1993) and Wulfert and Hayes (1988) in that symmetry was tested after a large number of transitivity tests and combined tests and, then, on test trials that were interpolated between combined tests and transitivity tests of many different nodal numbers. Even so, the results of the present study seem to diminish the importance of the test complexity variable, since reaction times to symmetry test trials were longer than to baseline trials in the absence of such a complex testing history and context.

Some authors have suggested that the "immediate emergence" of new classes is "the strongest index of equivalence class formation." (Fields et al., 1997; cf. also Sidman, 1992). By "immediately," these authors meant "the first presentation of all emergent relations probes." However, the reaction time measure permits a more fine-grained analysis of "immediacy." The finding of longer reaction times to comparison stimuli on test probes than on training trials, whether or not class consistent responding emerges, may be relevant to the question of *when* equivalence emerges. This question was originally asked because class consistent responding sometimes emerges only after repeated testing (cf. Sidman, Kirk, & Willson-Morris, 1985). The fact that reaction times typically increase initially during testing, even when class consistent responding occurs from the first presentation of each test probe, may suggest that those test probes might be perceived as "problems" that occasion some sort of precurrent responding before a correct or an incorrect selection occurs. For some subjects, the most significant problem may be developing a consistent pattern of responding to each type of test probe, whether or not this pattern is in accord with an experimenter-defined class. Thus, the 3 subjects who responded consistently to both symmetry, and to "equivalence" probes were also among those with the shortest overall reaction times. It is possible that responding in accord with equivalence in some experimental arrangements is the simplest way to achieve such consistency. In contrast, it is not obvious that inconsistent comparison selection constitutes a problem for all subjects. For instance, the only subject (#65) who responded as fast as those 3 who responded most consistently in the present experiment responded inconsistently to all trial types throughout most of the test phase. Thus, it is also possible that the facilitating effect of previously established equivalence classes (e.g., Buffington et al, 1997; Fields et al., 1997) is related to the pretraining of consistent comparison selection across trials rather than to the actual formation of classes.

Some experimenters (e.g., Barnes & Keenan, 1993; Dymond & Barnes, 1995; Roche & Barnes, 1996) have used a stability measure as a criterion for terminating testing. They applied the stability criterion to performance on two-choice tasks and characterized responding to trial types within a complete test block as high-rate or low-rate performance. The consistency index used in the present study, however, was the same as the one used by Holth and Arntzen (1998a), which characterized *changes* in the consistency of comparison selections on a three-choice task *throughout* the equivalence test phase. In that previous study (Holth & Arntzen 1998a), 19 of 25 subjects who did not respond in accord with equivalence still responded consistently in accord with some other pattern, whereas only 2 of 9 subjects did so in the present experiment. This difference is probably related to the different test procedures that were used. Whereas in the previous study, the separate equivalence test blocks included only three different trial types (i.e., either C1, C2, or C3 as a sample, and always A1, A2, and A3 as comparisons), in the test probe phase of the present experiment, those three trial types were mixed with six different training trials and six different symmetry test trials. Hence, the development of consistent responding may require more extended exposure to the current test probe procedure.

Although class consistent responding to all experimenter-defined classes occurred in only 1 of the 10 subjects, 5 of the remaining 9 subjects showed such class consistent responding with respect to a minimum of one of the three classes. Should these be considered as partial successes? For instance, if class consistent selection of A1 occurs reliably the presence of C1, although not the selection of A2 in the presence of C2, or A3 in the presence of C3, should we consider A1, B1, and C1 as related by equivalence, but not A2, B2, C2, and A3, B3, and C3, respectively? The problem with such an interpretation is that, given any consistent pattern of comparison selections, the probability of hitting upon a minimum of one of the three experimenter-defined classes equals 19/27 (cf. Appendix B). For instance, such class consistency could be obtained by choosing the same comparison for all three samples. Furthermore, consistency with respect to one class can be obtained without consistency with respect to the remaining two comparisons. Thus, the number of subjects that responded in accord with at least one experimenter-defined class in the present experiment hardly exceeded what was to be expected from pure chance.

The previous study (Arntzen & Holth, 1997) that showed low yields of equivalence in separate test blocks following the linear series design included only combined "equivalence" tests and symmetry tests and, therefore, only those tests were included in the present experiment. Although Fields, Adams, Newman, and Verhave (1992) used a training protocol that differed from the one in the present experiment, their results indicated that the emergence of transitivity may typically precede the emergence of class consistent responding on the combined test. Therefore, the inclusion of transitivity tests in addition to the symmetry tests might have increased the probability of responding in accord with equivalence. Also, the results of Fields et al. (1992) showed that class consistent responding emerged in some subjects following repeated testing, and Bush, Sidman, and de Rose (1989) found that following the establishment of symmetry, equivalence emerged only after a number of alternating symmetry and equivalence test blocks.

Experiment 2 investigated the effects of repeated training and testing with respect to a stable baseline performance, reaction times, consistency of responding during testing, and the probability of an equivalence outcome. Furthermore, if "equivalence" still did not emerge on the combined tests, transitivity tests would be included to investigate whether transitivity and "equivalence" would then emerge.

Experiment 2

Method

Subject

One subject (#61), who participated in Experiment 1, was asked to participate in a series of reexposures to the same training and test. The subject would then be exposed to the same training and test repeatedly until baseline performance continued undisturbed by the test probes to see if some systematic pattern of responding to the untrained relations then emerged.

Apparatus

The apparatus was the same as in Experiment 1.

Procedure

During eight reexposures (Sessions 2-9), the stimulus materials and basic procedure were the same as in Experiment 1, except that the introductory, separate AB and BC, training blocks were skipped during the last three exposures. Each session consisted of one training phase and one test phase—in which test trials were interspersed between training trials. In addition, the last three of these sessions (7-9) included a 24-trial transitivity test block prior to the usual test phase. Intersession intervals ranged from 1 to 3 days.

Table 3

Stimulus Materials Used During Final Sessions, 10 and 11 in Experiment 2

	4	5	6
A	β	η	σ
B	χ	α	ω
C	ε	υ	τ

Finally, two sessions (10 and 11) were run with a different set of stimulus materials, displayed in Table 3. A transitivity test was also included in the last session with the new materials.

Results

From the first reexposure (Session 2) and throughout the experiment, baseline performance was close to perfect, with number of errors during the 38 baseline trials in the test phases ranging only from zero to one. Responding in accord with symmetry was close to perfect in the fourth session and for the rest of the experiment. However, during nine

Table 4

Number of Training Trials to Criterion, Total Number of Errors During Training Blocks, and Indices of Responding in Accord with Baseline, Symmetry, "Equivalence," and Transitivity During Repeated Training and Testing with Subject #61, Experiment 2

	Training Phase			Baseline training trials				Test Phase Symmetry probes		Equivalence probes		Trans-itivity
Sess. No.	Stim. Classes	No. of Trials	No. of Errors	1-10	11-20	21-30	31-38	1-10	11-20	1-10	11-20	24-trial Block
1	1-3	96	2	**9/10**	8/10	8/10	6/8	**9/10**	**9/10**	7/10	7/10	-
2		130	3	**10/10**	**10/10**	**10/10**	7/8	7/10	**9/10**	7/10	6/10	-
3		96	2	**10/10**	**10/10**	**10/10**	**8/8**	**9/10**	7/10	4/10	8/10	-
4		101	1	**10/10**	**10/10**	**10/10**	**8/8**	**10/10**	**9/10**	3/10	6/10	-
5		106	1	**9/10**	**10/10**	**10/10**	**8/8**	**9/10**	**10/10**	3/10	6/10	-
6		98	1	**10/10**	**10/10**	**10/10**	**8/8**	**9/10**	**10/10**	4/10	5/10	-
7		27	1	**10/10**	**10/10**	**10/10**	**8/8**	**10/10**	**10/10**	4/10	6/10	3/24
8		24	0	**9/10**	**10/10**	**10/10**	**8/8**	**9/10**	**10/10**	5/10	5/10	12/24
9		24	0	**10/10**	**10/10**	**10/10**	**8/8**	**10/10**	**9/10**	3/10	5/10	9/24
10	4-6	113	4	**10/10**	**10/10**	**10/10**	**8/8**	**9/10**	**9/10**	5/10	3/10	-
11		103	2	**10/10**	**10/10**	**9/10**	**8/8**	**10/10**	**9/10**	5/10	8/10	16/24

Note. Session 1 was part of Experiment 1.

exposures to the same training and test conditions, summing up to a total of 1010 training trials, 180 symmetry test trials, 180 equivalence test trials, and 72 transitivity test trials, the subject still did not respond in accord with equivalence (see Table 4). During the transitivity test blocks in Sessions 7-9, the subject did not respond in accord with transitivity.

Then, with the new stimulus materials in Sessions 10 and 11, baseline performance was not even disrupted initially during testing, and symmetry performances were almost perfect from the start of testing. Yet, the subject did not respond in accord with equivalence. As in the tests with the previous stimulus materials in Sessions 7-9, a pattern of responding that was not in accord with transitivity emerged during the transitivity test block with the second set of stimulus materials in Session 11.

However, other patterns of consistent responding to the "equivalence" test trials gradually emerged with both sets of stimulus materials. The patterns of consistent responding displayed by the subject during the last session with the first set of materials (Session 9) are illustrated in Figure 4. All 38 baseline trials were correctly completed, and 19 of 20 symmetry

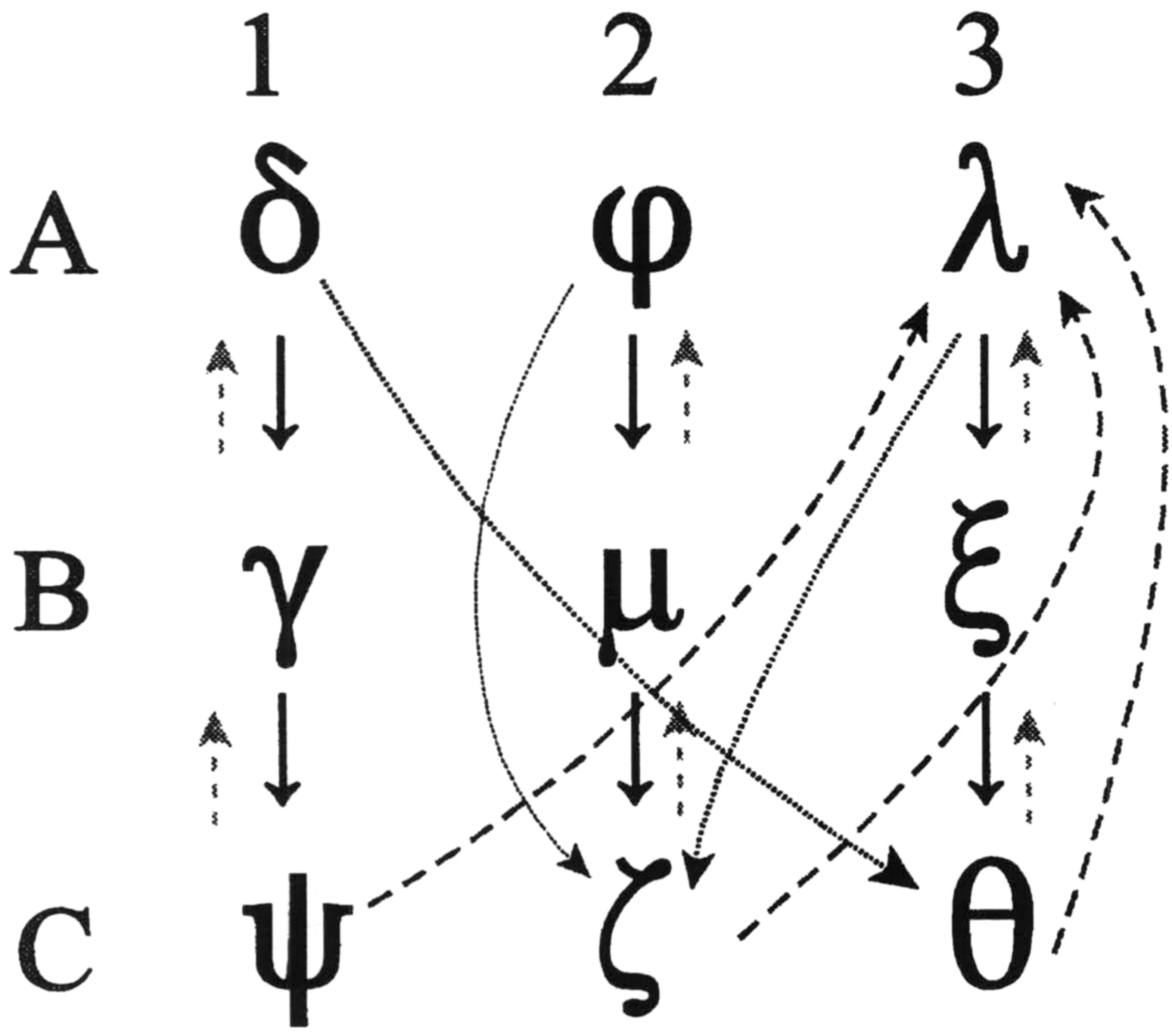

Figure 4. Patterns of consistent responding for Subject #61 during the last session (9) with the first stimulus materials (Classes 1-3). Solid arrows from A to B and from B to C show responding in accord with the trained relations. Arrows from C to B and from B to A show the pattern of responding to symmetry test probes, arrows from C to A indicate the pattern of responding during equivalence probes and, finally, arrows from A to C illustrate the pattern of selection during the transitivity test block.

test probes were responded to in accord with symmetry. On equivalence test probes, the subject chose comparison A3 in the presence of each of the three samples, C1, C2, and C3. During the transitivity test block, the subject chose the comparison C3 when A1 was the sample, and C2 when either A2 or A3 was the sample.

The corresponding patterns of consistent responding during the final session (11) with the second set of materials are shown in Figure 5. Of 38 baseline trials, 37 were correctly completed, 19 of 20 test probes were completed in accord with symmetry. On CA equivalence test probes, the subject chose A1 in the presence of C1, but A3 in the presence of both C2 and in the presence of C3. During AC transitivity test probes, the subject chose C1 in the presence of A1, but chose C2 in the presence of both A2 and A3.

Responding throughout each session was assessed with respect to whether it was in accord with the known finally emerging pattern, and/or whether it was in accord with the experimenter-defined equivalence pattern. Figure 6 shows that responding in accord with the final pattern emerged already in the second session, as indicated by the solid line

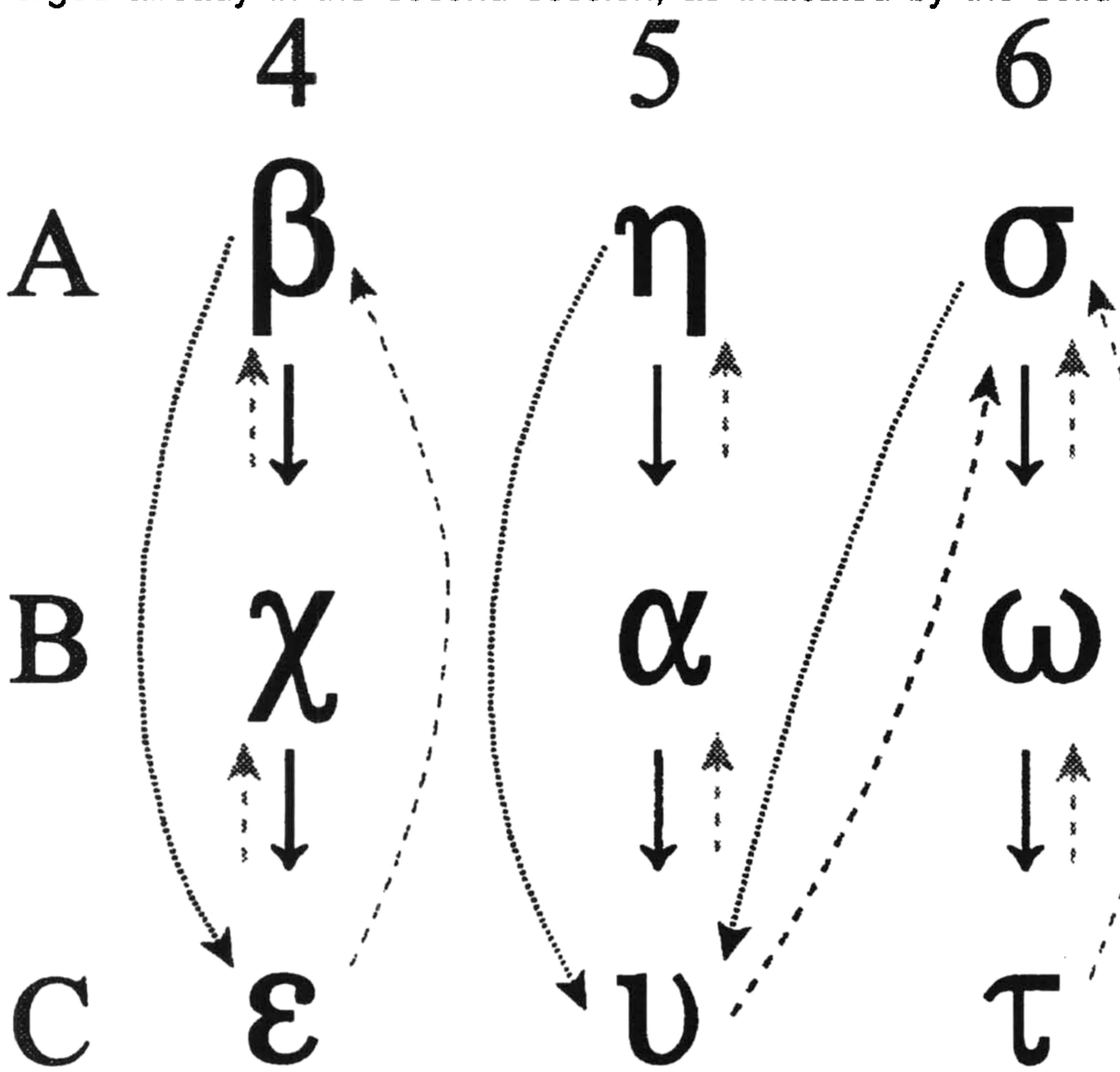

Figure 5. Patterns of consistent responding for Subject #61 during the final session (11) with the second set of stimulus materials (Classes 4-6). Arrows indicate patterns of responding to trained relations and to different test trials as in Figure 4.

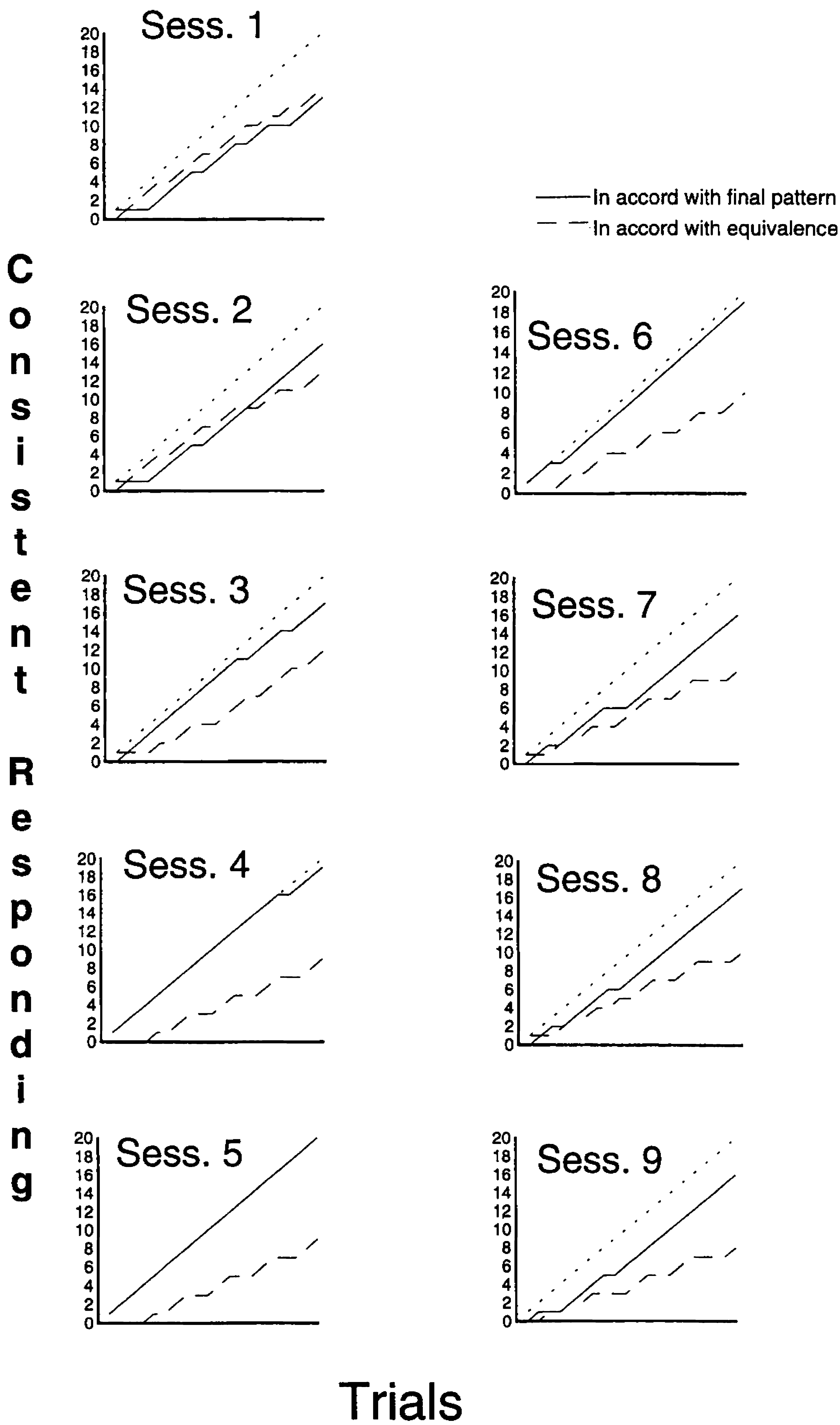

Figure 6. Each panel of the figure shows, cumulatively, comparison stimulus selections in accord with the finally emergent pattern (solid line), and selections in accord with experimenter-defined equivalence classes (dashed line) during "equivalence" test trials. Any deviance from the slope of the dotted diagonal illustrates the degree of discordance with either of the respective patterns.

being parallel with the dotted diagonal which shows the slope of the curve in the case of complete consistency with either pattern. With the exception of some minor inconsistency towards the end of Sessions 3

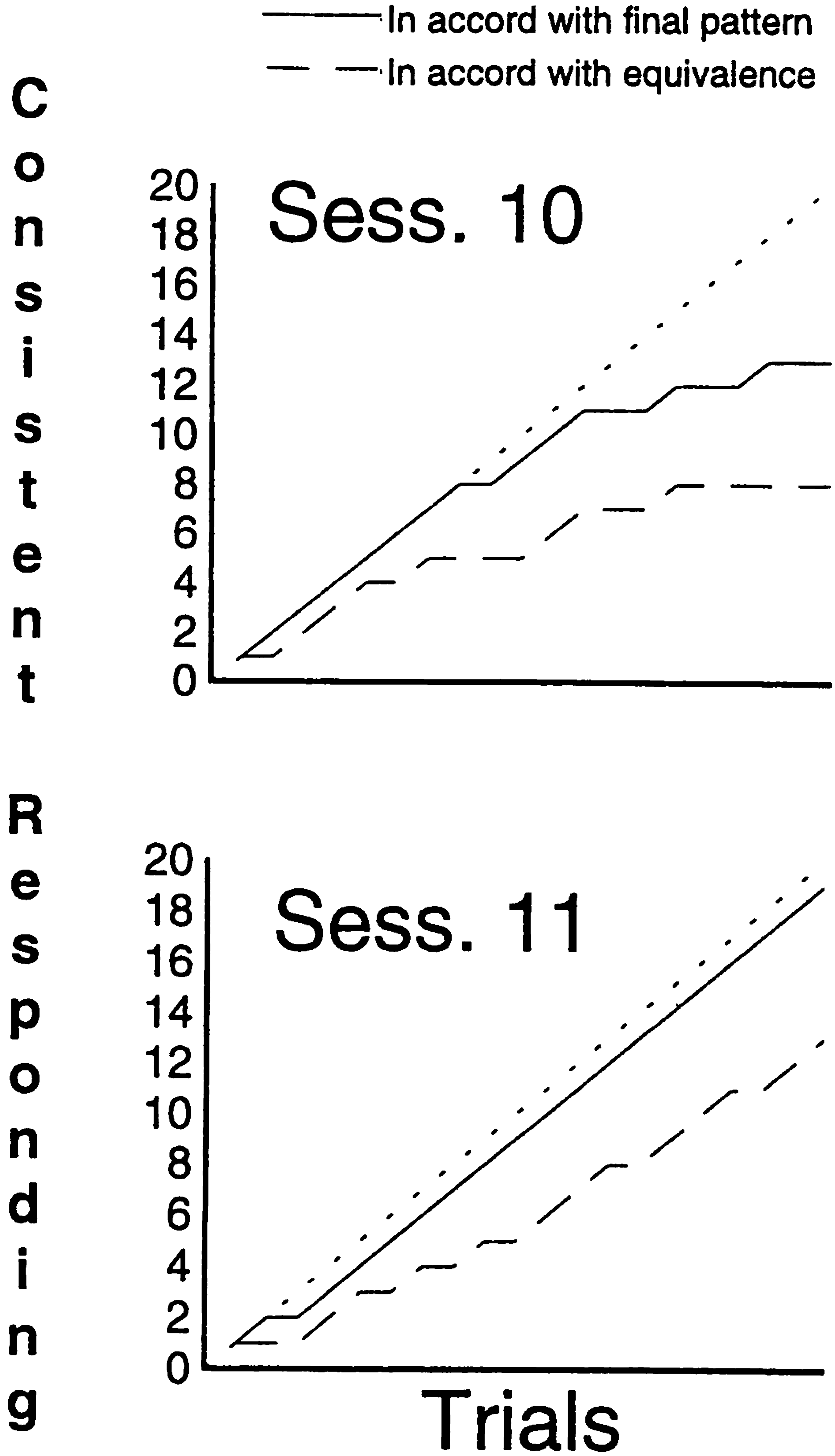

Figure 7. As in Figure 6, the solid line shows, cumulatively, selections in accord with the final pattern, and the dashed line shows selections in accord with experimenter-defined equivalence.

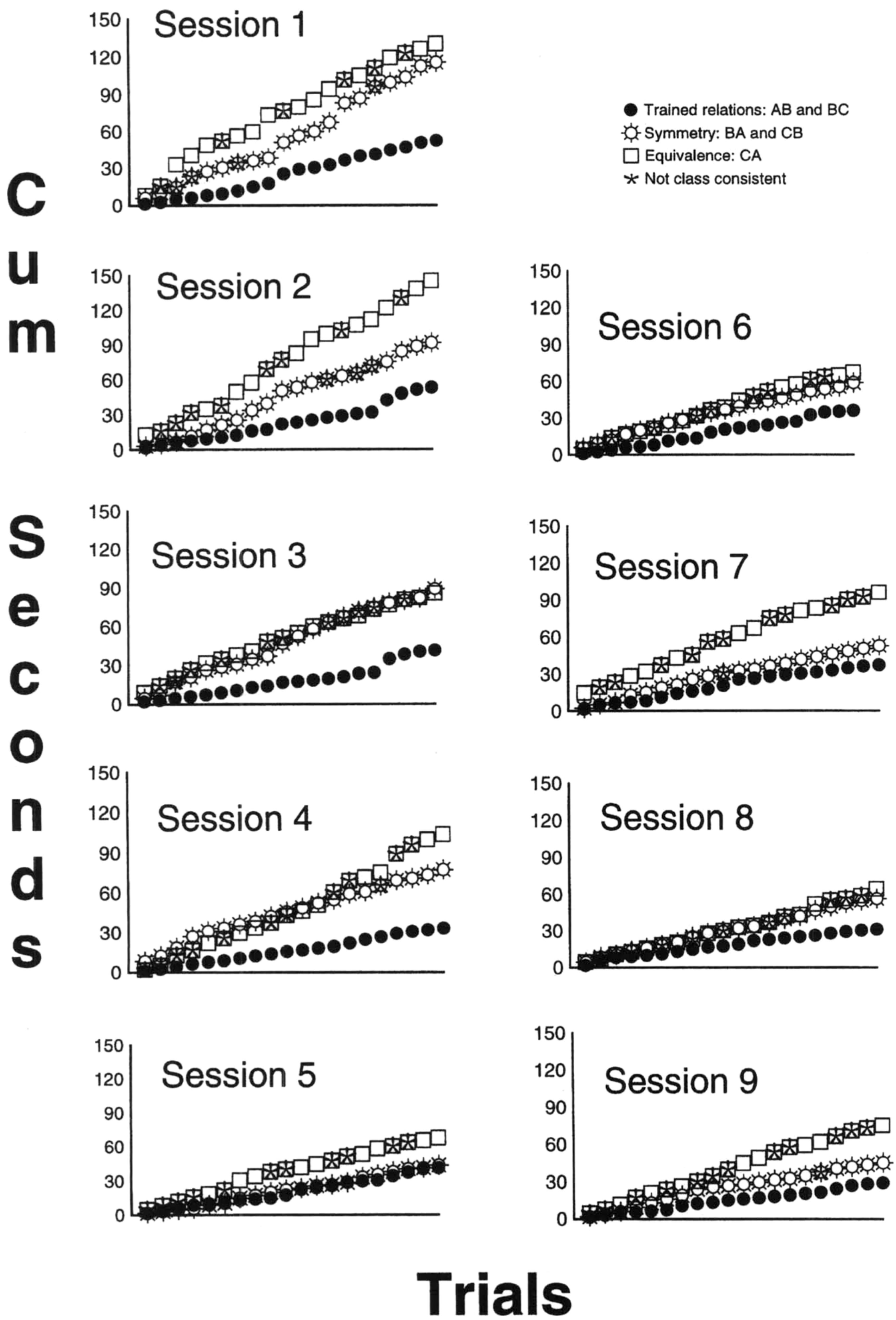

Figure 8. Cumulative reaction times to comparison stimuli during equivalence test probes, symmetry test probes, and the first 20 training trials for Subject #61 in each of 9 successive sessions with the first set of stimulus material in Experiment 2. (The first session is from Experiment 1).

and 4, and early in Sessions 6-9, responding remained consistent with this pattern throughout all sessions with the first set of materials. The dashed line in successive panels of Figure 6 shows that gradually fewer selections were made in accord with equivalence over the first few sessions, ending with approximately 1/3 of the selections being made in accord with the equivalence pattern.

Responding in accord with equivalence and responding with the final pattern with the second set of stimulus materials (Sessions 10 and 11) are shown in Figure 7. No consistent pattern emerged during the first session (10) with the new material. As shown by the solid line being parallel with the consistency diagonal in the lower panel of the Figure 7, responding was

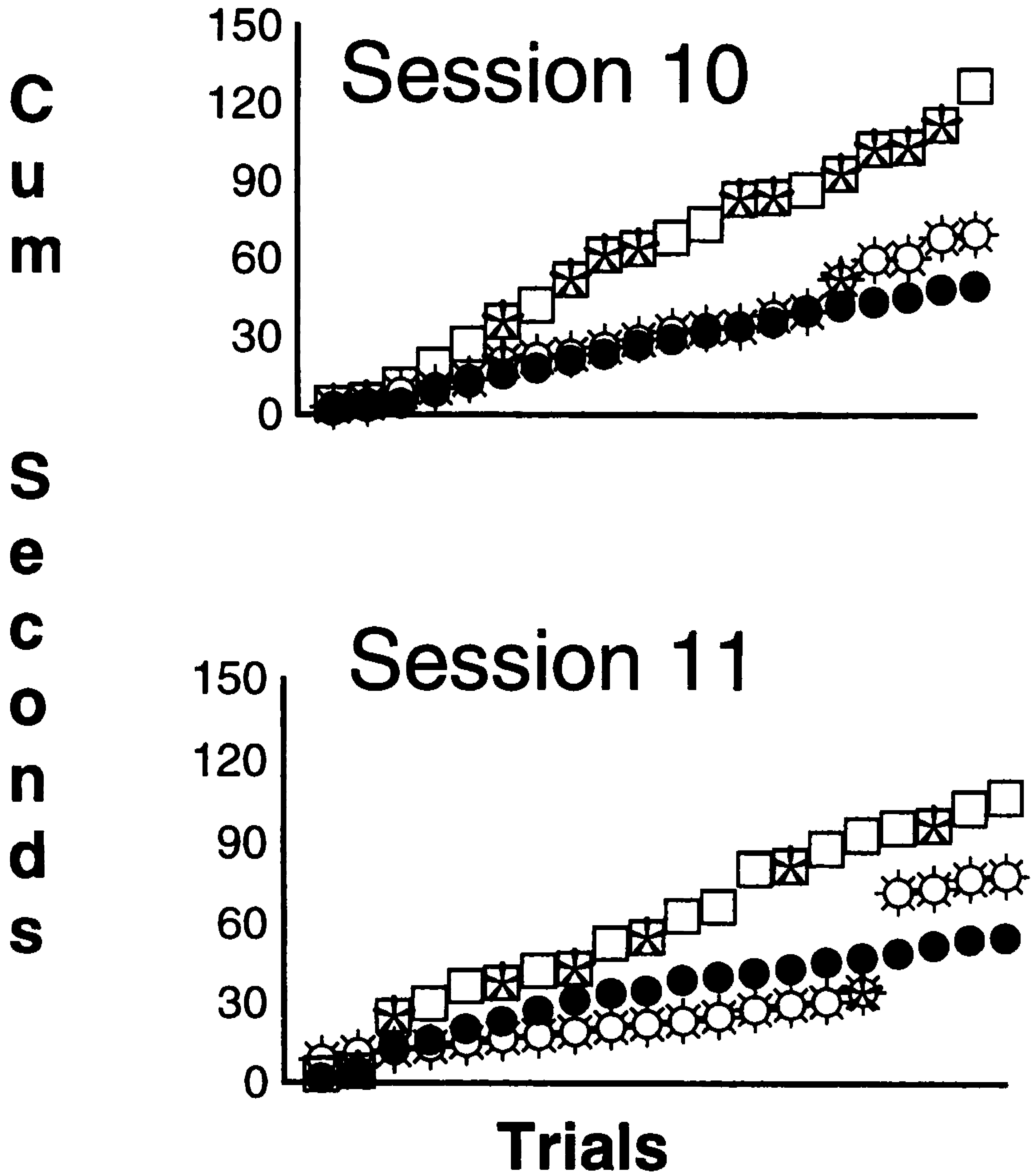

Figure 9. Cumulative reaction times to comparison stimuli during equivalence test probes, symmetry test probes, and the first 20 training trials for Subject #61 in each of 9 successive sessions with the second set of stimulus material in Experiment 2.

consistent throughout most of the second session (11) with the new materials. This pattern implied that approximately 2/3 of the selections during the "equivalence" test were made in accord with equivalence.

Cumulative records of reaction times to comparison stimuli during baseline trials, symmetry probes, and equivalence probes in Sessions 1-9 are displayed in Figure 8. The corresponding curves for reaction times with the new materials in Sessions 10 and 11 are shown in Figure 9.

Reaction times to all trial types were gradually lowered as a function of repeated training and testing. In all sessions except for one (Session 5), reaction times to test probes were markedly longer than to baseline trials. Furthermore, during five of these nine sessions, reaction times to equivalence probes were substantially longer than to symmetry probes. The reverse was never the case. With the new materials in Sessions 10 and 11, responding to equivalence probes was slower than to symmetry and to baseline trials. Except for one trial in each of these two sessions, responding to symmetry probes was as fast as, or faster than, responding to baseline trials.

Discussion

The results showed that even after repeated training and testing, when baseline as well as symmetry performances are nearly perfect, responding during "equivalence" and transitivity testing can still develop into patterns that are discordant with experimenter-defined equivalence. The likelihood of this finding being caused by peculiar characteristics of the stimulus materials is substantially lowered by the replication with a second set of materials during the latter two sessions of the present experiment.

Previous studies (e.g., Buffington et al., 1997; Fields et al., 1997) have demonstrated that the probability of forming new classes in a simultaneous testing protocol can be substantially increased as a function of previously successfully established classes. The results of the present experiment indicate that repeated training and testing that do not result in class formation may not increase the likelihood of forming new classes.

The results of Experiment 2 extend those of Holth and Arntzen (1998a) by showing that in spite of a prolonged training that produced a perfect baseline performance throughout testing, and perfect symmetry, responding during "equivalence" and transitivity test probes can still occur in discordance with the experimenter-defined classes. The lack of responding in accord with transitivity when symmetry is established, and "equivalence" is not, is in accord with previous results obtained by Fields and coworkers (1992). They suggested that if symmetry performances are class consistent and performances on the abbreviated equivalence test are not, transitivity cannot be intact. Yet, responding consistently in accord with a different pattern emerged during both the "equivalence" test and during the transitivity test with both sets of stimulus materials.

Harrison and Green (1990) showed that a pattern of consistent responding during testing could develop independently of baseline performances when the test procedure included comparisons among which only one was invariably present for each sample throughout testing. However, no such multiple negative comparison training was involved in the present experiment. An alternative source of the consistent selection of a single comparison for each sample reported here may be the simple fact that advancement through training had always required such consistency. Thus, inconsistent comparison selection may become a major 'problem' to be solved during "equivalence" testing. In accord with this interpretation, reaction times decreased over repeated training and testing, as responding became gradually solidified into the final pattern. If some type of precurrent responding is a facilitating variable with respect to class formation, it would also follow the prevention of longer reaction times initially during training might disrupt class formation even when using an otherwise more effective training protocol.

Experiment 3

The purpose of the third experiment was to investigate the effect upon the probability of stimulus equivalence when reaction time constraints are placed upon responding to comparison stimuli during testing. Hence, baseline conditional discrimination training was conducted according to a one-to-many (often called sample-as-node) design that in a previous study by Arntzen and Holth (1997) was shown to result in a high probability of stimulus equivalence. In that study, all 10 of 10 subjects responded in accord with stimulus equivalence following the one-to-many training. Mean reaction times to comparison stimuli during the final stage of training were approximately 2 s, while the reaction times initially during equivalence testing ranged from approximately 5-12 s. Hence, specifically, the present experiment was concerned with whether class consistent responding will emerge following one-to-many training when reaction times are constrained to an interval that is longer than the typical reaction times towards the end of baseline training, but substantially shorter than what is commonly observed initially during the test.

Method

Subjects

Ten psychology students served as subjects. None of the subjects had previously participated in experiments on stimulus equivalence.

Apparatus

The apparatus and the stimulus materials were the same as in the previous experiments.

Procedure

General information and instruction. In addition to the information and instruction given in Experiment 1, the following instruction was given by the experimenter: "During some parts of the experiment, you will have only two seconds available for choosing a comparison stimulus on the right side of the monitor."

Training and test. The basic training procedure was the same as in Experiment 1, with the following exceptions: First, the training was completed according to a one-to-many design, BA and BC. Second, during the final stage of training, with 24 randomly intermixed BA and BC trials, the mastery criterion required successive correct completion of all 24 trials with reaction times to comparison stimuli never exceeding 2.0 s. When a subject pressed an incorrect key or failed to respond to comparison stimuli within the specified 2-s interval, another 24 correctly completed trials were required before the subject could advance to the first equivalence test. Third, if a subject had not completed the tasks after 65 min, the experimenter would enter the room to terminate the session.

The training was followed by two 24-trial CA equivalence tests. During the first test, reaction times to comparison stimuli were constrained to 2 s as in the final part of training. A 2-s intertrial interval was initiated whenever the subject selected a stimulus within the limits of the reaction time requirement, or when the subject failed to meet that requirement. Pressing a key during the intertrial interval had no programmed consequences. During the second 24-trial CA equivalence test, there were no reaction time requirements.

Results

Of the 10 subjects, 5 (#901, #902, #903, #906, and #908) completed training. Because of a subtle programming error, 2 of these subjects advanced from training to test without fully meeting the strict training criterion. Subjects #901 and #906 responded correctly and within the specified reaction time to 19 and 21, respectively, of the last 24 trials during training.

Table 5

Number of Training Trials, Errors, and Trials on which Reaction Time Criterion Was Not Met During Baseline Training, Number of Trials Completed in Accord with Equivalence, Number of 'Errors,' and Number of Trials on which Reaction Time Requirement Was Not Met During Testing

Training				CA Equivalence Time Restriction: 2.0 s						CA Equiv NoTime Restriction	
				1-12			13-24			25-36	37-48
Subj No.	No. of Trials	No. of Errors	Time Out	Eq	Err	Time Out	Eq	Err	Time Out	Eq	Eq
903	202	6	5	5	1	6	5	3	4	**12**	**12**
906	447	30	11	4	0	8	4	1	7	10	**11**
901	259	11	6	6	1	5	3	5	4	8	**11**
908	480	28	14	0	3	9	7	4	1	5	1
902	341	12	12	2	6	4	1	8	3	4	3
904	744	25	35	-	-	-	-	-	-	-	-
905	737	14	38	-	-	-	-	-	-	-	-
907	621	17	24	-	-	-	-	-	-	-	-
909	600	23	27	-	-	-	-	-	-	-	-
910	574	27	21	-	-	-	-	-	-	-	-

Note. The last two columns show the number of trials that were completed in accord with equivalence during each half of a final 24-trial test block when there was no reaction time requirement.

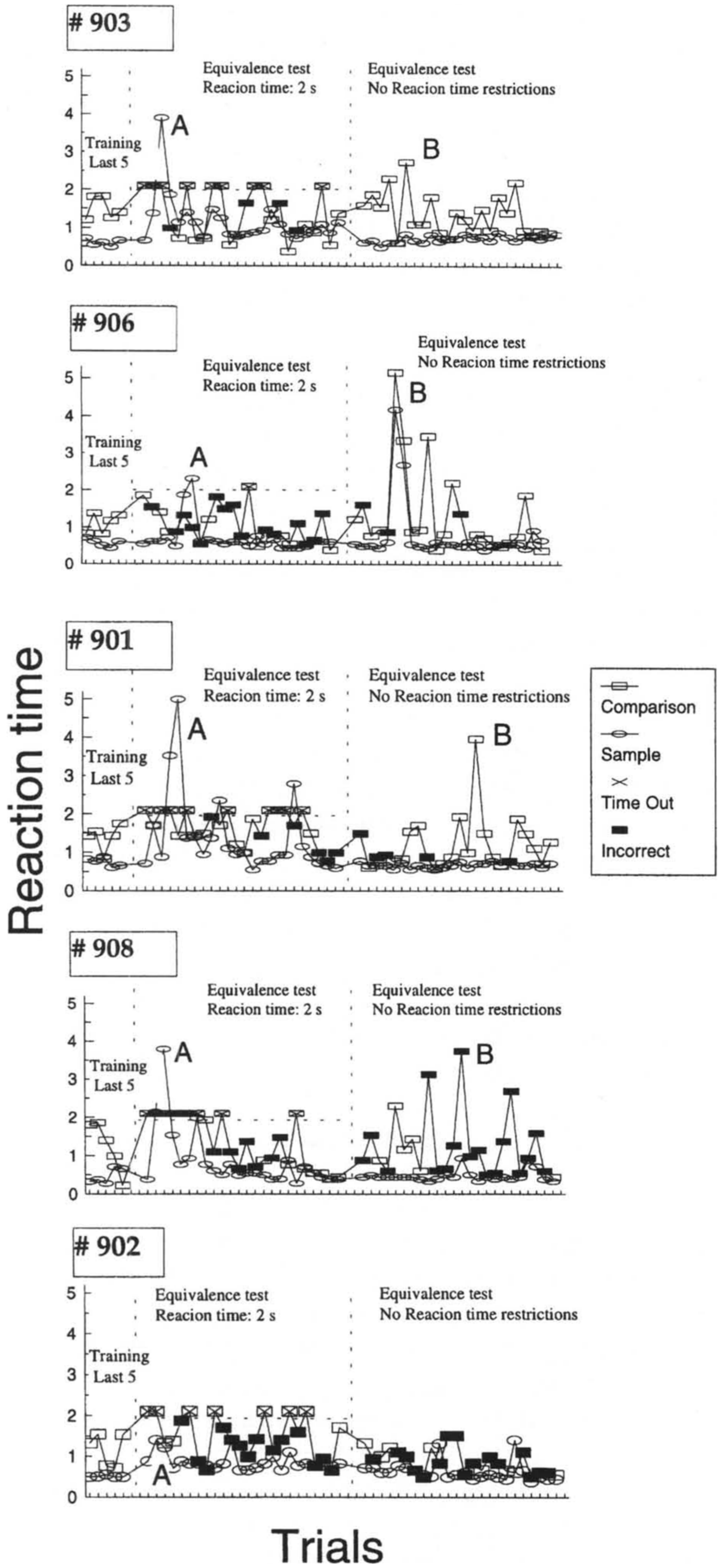

Figure 10. Each panel shows individual reaction times to samples and comparisons in Experiment 3 during the last five training trials, tests trials with a 2-s reaction time criterion and, finally, when the reaction times criterion was discontinued.

During the first equivalence test (with the 2-s reaction time requirement for choosing comparison stimuli), none of the 5 subjects reliably responded in accord with equivalence. Individual data for these 5 subjects are presented in Table 5. For each half of the test, the table shows the number of responses in accord with experimenter-defined equivalence, the number of responses in discordance with equivalence, and the number of trials on which responding to comparison stimuli failed to meet the reaction time requirement. For each subject, there was a higher number of failures to meet the reaction time criterion during the first test half, while the number of responses in discordance with experimenter-defined equivalence was higher in the second test half.

During the second equivalence test, where there were no reaction time requirements, 1 subject (#903) responded consistently in accord with equivalence in the first test half, while 2 more subjects (#901 and #906) responded in accord with equivalence during the second test half.

For all subjects, reaction times to sample stimuli were stable at approximately 0.5 s during training and during most of the test. However, when there was a reaction time requirement to comparison stimuli, reaction times to samples increased (marked with an A in Figure 10).

In the second test, when there was no longer a reaction time requirement, 3 of the 5 subjects (#901, #903, and # 906) responded in accord with equivalence. For these 3 subjects, reaction times to comparison stimuli were at some point (marked with B in Figure 10) longer than the previous requirement. The fourth subject (#908) who also showed longer reaction times during the second equivalence test did not respond in accord with equivalence, but responded consistently in accord with a different pattern after those trials on which reaction times were longer. The only subject (#902) who did not show any such increased reaction time to comparisons during testing responded consistently, although not in accord with equivalence, from the latter part of the first test and throughout the second.

Discussion

Experiment 3 was concerned with the likelihood of stimulus equivalence formation when reaction times were kept within limits typically observed during the final parts of baseline training. The reaction time limit of 2 s was also chosen on the basis of previous experimental results. With the same apparatus, Arntzen and Holth (1997) and Holth and Arntzen (1998a) found that reaction times to comparison stimuli were typically below this value both during the final parts of training and during the later parts of tests, while reaction times initially during tests for emergent relations were typically substantially above the 2-s value. The zero-of-five yields of an equivalence outcome in the present experiment when only 2 s were available for responding to comparison stimuli, and the three-of-five yields after the reaction time requirement was lifted, are dramatically lower than the ten-of-ten yields obtained with the same apparatus, training and testing procedures, and stimulus materials with no reaction time constriction (Arntzen & Holth, 1997). Thus, in the absence of an opportunity for longer reaction times initially during testing, the emergence of stimulus equivalence can be seriously impeded.

General Discussion

Using a simultaneous testing protocol following linear series training provides a framework for investigating effects of variables that are not apparent using more effective procedures. In addition, the reaction times data support the conclusion drawn by Spencer and Chase (1996) that this measure appears to be sensitive to different variables even when accuracy of responding is not. Reaction times during testing are not only sensitive to other variables. The lack of responding in accord with equivalence in the presence of the reaction time constriction in Experiment 3 suggests that events associated with those longer reaction times can play a very significant role in the formation of class consistent responding.

Generally, the literature on stimulus equivalence has not indicated that equivalence can often fail to emerge even on some conventional procedures. Rather, it has been suggested that the descriptive/functional discrepancies that characterize equivalence may be considered as an unanalyzable primitive (e.g., Sidman, 1990; Stoddard & McIlvane, 1986). Three somewhat different lines of argument seem to have led to this currently dominant view that stimulus equivalence may be considered as a fundamental stimulus function. First, after considering some evidence against a particular form of mediation by common naming of the relevant sample-comparison pairs, Sidman, Willson-Morris, and Kirk (1986) established the view that no form of mediation needs to be involved and, hence, that stimulus equivalence may adequately be considered as resulting directly from conditional discrimination training. Second, the apparently high generality of the phenomenon has led to the interpretation of occasional failures to confirm to the predicted equivalence outcomes as artifactual (e.g., Sidman, 1993), and a number of artifactual results have actually been investigated (e.g., Sidman, 1993, 1994). Third, the rapidity with which equivalence emerges has been taken to indicate that stimulus equivalence must be attributed to "uncontrolled variables," suggesting simply that "we are made that way" (Sidman, 1992, p. 23).

In order to truly delineate the generality of equivalence phenomena, however, research pertinent to conditions under which the predicted emergent relations do not emerge is also needed. The present series of experiments indicate that (1) Even a perfectly maintained baseline with respect to the prerequisite conditional discriminations does not ensure the emergence of experimenter-predicted emergent performances in accord with equivalence; (2) responding may still gradually become consistent with some other pattern; (3) reaction times are typically longer initially during tests for emergent relations; and (4) when those longer reaction times initially during testing are prevented, emergence of the predicted emergent relations is seriously impeded.

Thus, the present research questions all of the above mentioned lines of argument for accepting stimulus equivalence as a basic process. First, stimulus equivalence often does not seem to emerge immediately, at least if the reaction time measure is brought into the definition of "immediacy." Second, following the prototypical AB, BC training, equivalence often does not seem to emerge at all. Finally, potential sources of mediation other than the common naming of sample-comparison pairs remain to be considered in more detail.

In a very persuasive account of novel performances, Sidman (1986) started by considering patterns of discrepancies between descriptive classes and functional classes that involve untrained relations of high generality, resulting from the well-established two- and three-term analytic units. From there, he extended the analysis to four- and five-term units that were also stripped of any kinds of mediational processes even in the account of complex contextual control. As an alternative, we suggest that the analysis might just as properly start by considering patterns of emergent relations that involve untrained relations of somewhat less generality. A case in point is 'problem solving,' in which mediating events are more obviously involved.

Palmer (1991) distinguished between "Memory as a stimulus control phenomenon" and "Memory as a problem solving phenomenon." Consider, for an average math-educated person, what makes answering "7 x 8 = ___" different from answering "75 x 85 = ___"? Presumably, for most persons with some math education, the first task simply serves as a verbal discriminative stimulus that occasions the answer as an intraverbal response. In contrast, the latter task requires some kind of precurrent responding, that is, problem solving. As pointed out by Catania (1992), when math problems are solved by paper and pencil, the intermediate products obviously serve as discriminative stimuli that may occasion the solution. Furthermore, "[p]resumably, the intermediate products would still enter into the solution, even if there was no written record of them. If we did not say them aloud, an observer might say we had engaged in 'mental arithmetic' . . . . But the role of the intermediate products is the same even if they are more public and more permanent in the first case than in the second. We still have much to learn about such processes, but we need not treat them as something other than behavior. (Catania, 1992, p. 348).

The question raised here is whether subjects on the first few equivalence test trials select the correct comparison directly under control of the sample. Alternatively, do correct comparison selections require precurrent responses to produce intermediate products that constitute discriminative stimuli for comparison selections? The absence of responding in accord with stimulus equivalence during the test with a 2-s reaction time restriction strongly suggests that correct comparison selection is not directly controlled by the sample initially during testing, and that an opportunity for precurrent responding may be relevant. Rather than simply accepting that "mental events take time" (e.g., Ashcraft, 1989), a behavior analysis might be concerned with what happens in terms of behavior when reaction times increase. In particular, reaction times initially during testing might be related to what has been called "acquisition and remembering strategies" (e.g., Delaney & Austin, 1998; Donahoe & Palmer, 1994; Palmer, 1991). As Sidman (1994) noted: "Any description of equivalence phenomena must not only be internally consistent but must also fit into that more general framework of data and principles" (p. 527). There is a literature on "acquisition and remembering strategies" both from the study of "memory experts" and from "laboratory learning tasks," with data that indicate strongly that precurrent responding both during training and test are relevant to a number of test performances in human subjects (see Delaney & Austin, 1998, for a review; cf. also Catania, 1992). The notion of precurrent responding may even be relevant to the finding of a higher probability of equivalence following one-to-many training as compared with the probability of equivalence following linear training. That is, one-to-many training may well foster a pattern of precurrent responding that is different from a pattern produced by linear training, because only the one-to-many design provides direct training in relating each sample to more than one comparison stimulus. In contrast, during linear training, each sample is related to only one comparison, and the baseline training could thus produce "looking for" the correct comparison even prior to the actual presentation of comparison stimuli.

Behavior analysts have only just begun to analyze in detail the complexity of such repertoires of precurrent responses that may be involved in the emergence of stimulus equivalence (e.g., Horne & Lowe, 1996; Lowenkron, 1998). Even if it should eventually turn out unequivocally that equivalence does not always require any form of precurrent responding, it is difficult to imagine that such phenomena are irrelevant to equivalence formation in humans (cf. Eikeseth & Smith, 1992). A variety of studies using protocol analyses with verbal subjects (e.g., Austin & Delaney, 1998; Wulfert et al., 1991), and establishing different overt precurrent responses in nonverbal organisms (e.g., Lowenkron, 1991, 1996; McIntire, Cleary, & Thompson, 1987) may be required to complete the story on how equivalence relations emerge. It is our view that an analysis that includes histories responsible for precurrent or problem solving "strategies" in which human subjects may typically engage will yield a more powerful tool in predicting and controlling the emergence of stimulus equivalence as well as consistent nonequivalence in human subjects.

References

ADAMS, B. J., FIELDS, L., & VERHAVE, T. (1993). Effects of test order on intersubject variability during equivalence class formation. *The Psychological Record*, 43, 133-152.

ARNTZEN, E., & HOLTH, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record*, 47, 309-320.

ASHCRAFT, M. H. (1989). *Human memory and cognition.* New York: HarperCollins.

AUSTIN, J., & DELANEY, P. F. (1998). Protocol analysis as a tool for behavior analysis. *The Analysis of Verbal Behavior*, 15, 41-56.

BARNES, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record*, 44, 91-124.

BARNES, D., & KEENAN, M. (1993). A transfer of functions through derived arbitrary and nonarbitrary stimulus relations. *Journal of the Experimental Analysis of Behavior*, 59, 61-81.

BENTALL, R. P., DICKINS, D. W., & FOX, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology*, 46B, 187-214.

BUFFINGTON, D. M., FIELDS, L., & ADAMS, B. J. (1997). Enhancing equivalence class formation by pretraining of other equivalence classes. *The Psychological Record*, 47, 69-96.

BUSH, K. M., SIDMAN, M., & DE ROSE, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior*, 51, 29-45.

CATANIA, A. C. (1992). *Learning* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

DELANEY, P. F., & AUSTIN, J. (1998). Memory as behavior: The importance of acquisition and remembering strategies. *The Analysis of Verbal Behavior*, 15, 75-91.

DONAHOE, J. W., & PALMER, D. C. (1994). *Learning and complex behavior.* Boston, MA: Allyn & Bacon.

DYMOND, S., & BARNES, D. (1995). A transformation of self-discrimination response function in accordance with the arbitrary applicable relations of sameness, more than, and less than. *Journal of the Experimental Analysis of Behavior*, 64, 163-184.

EIKESETH, S., & SMITH, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: The role of naming. *Journal of the Experimental Analysis of Behavior*, 58, 123-133.

FIELDS, L., ADAMS, B. J., NEWMAN, S., & VERHAVE, T. (1992). Interactions among emergent relations during equivalence class formation. *The Quarterly Journal of Experimental Psychology*, 45B, 125-138.

FIELDS, L., ADAMS, B. J., VERHAVE, T., & NEWMAN, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior*, 53, 345-358.

FIELDS, L., LANDON-JIMENEZ, D. V., BUFFINGTON, D. M., & ADAMS, B. J. (1995). Maintained nodal distance effects after equivalence class formation. *Journal of the Experimental Analysis of Behavior*, 64, 129-146.

FIELDS, L., REEVE, K. F., ROSEN, D., VARELAS, A., ADAMS, B. J., BELANICH, J., & HOBBIE, S. A. (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.

HARRISON, R. J., & GREEN, G. (1990). Development of conditional equivalence relations without differential consequences. *Journal of the Experimental Analysis of Behavior*, 54, 225- 237.

HOLTH, P., & ARNTZEN, E. (1998a). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record*, 48, 81-110.

HOLTH, P., & ARNTZEN, E. (1998b). Symmetry versus sequentiality related to prior training, sequential dependency of stimuli, and verbal labeling. *The Psychological Record*, 48, 293-315.
HORNE, P. J., & LOWE, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65, 181-241.
LOWENKRON, B. (1991). Joint control and the generalization of selection-based verbal behavior. *The Analysis of Verbal Behavior*, 9, 121-126.
LOWENKRON, B. (1996). Joint control and word-object bidirectionality. *Journal of the Experimental Analysis of Behavior*, 65, 252-255.
LOWENKRON, B. (1998). Some logical functions of joint control. *Journal of the Experimental Analysis of Behavior*, 69, 327-354.
MANDELL, C., & SHEEN, V. (1994). Equivalence class formation as a function of the pronounceability of the sample stimulus. *Behavioural Processes*, 32, 29-46.
MCINTIRE, K. D., CLEARY, J., & THOMPSON, T. (1987). Conditional relations by monkeys: Reflexivity, symmetry, and transitivity. *Journal of the Experimental Analysis of Behavior*, 47, 279-285.
PALMER, D. C. (1991). A behavioral interpretation of memory. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior: The first international institute on verbal relations* (pp. 261-279). Reno, NV: Context Press.
ROCHE, B., & BARNES, D. (1996). Arbitrarily applicable relational responding and sexual categorization: A critical test of the derived difference relation. *The Psychological Record*, 46, 451-475.
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.
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.
SIDMAN, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213-245). Hillsdale, NJ: Erlbaum.
SIDMAN, M. (1987). Two choices are not enough. *Behavior Analysis*, 22, 11-18.
SIDMAN, M. (1990). Equivalence relations: Where do they come from? In D. E. Blackman & H. Lejune (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 93-114). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
SIDMAN, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 15-27). Reno, NV: Context Press.
SIDMAN, M. (1993). Stimulus equivalence in and out of the laboratory. In A. Brekstad & G. Svedsäter (Eds.), *Proceedings from the 21st Annual Congress of the European Association for Behaviour Therapy, in Oslo, Norway, September, 1991*. Slependen, Norway: Norwegian Association for Behavior Analysis.
SIDMAN, M. (1994). *Equivalence relations and behavior: A research story.* Boston: Authors Cooperative.
SIDMAN, M. (1997). Equivalence: A theoretical or a descriptive model? *Mexican Journal of Behavior Analysis*, 23, 125-145.
SIDMAN, M., KIRK, B., & WILLSON-MORRIS, M. (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 discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37, 5-22.

SIDMAN, M., WILLSON-MORRIS, M., & KIRK, B. (1986). Matching-to-sample prodedures and the development of equivalence relations: The role of naming. *Analysis and Intervention in Developmental Disabilities*, 6, 1-19.

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, 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.

STIKELEATHER, G., & SIDMAN, M. (1990). An instance of spurious equivalence relations. *The Analysis of Verbal Behavior*, 8, 1-11.

STODDARD, L. T., & MCILVANE, W. J. (1986). Stimulus control research and developmentally disabled individuals. *Analysis and Intervention in Developmental Disabilities*, 6, 155-178.

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.

Appendix A

Instruction

When a subject was seated in front of the monitor, text on the sample stimulus key said: "Press here when you are ready to start." The experimenter then gave the following instructions: "When you touch the stimulus on the left side of the screen, one or more stimuli will appear on the right side. One of these is the correct stimulus for this trial. A touch on the correct stimulus will be followed by music from the cassette player, while an incorrect selection will be followed by the blanking of the screen for 5 s before a stimulus in the left-hand key is presented again. Each part of the training requires a certain number of correct responses before proceeding to the next part. The training will be followed by tests, in which there will be no differential consequences for correct and incorrect responses - no music and no blank screen."

Appendix B

All Possible Patterns of Consistent Comparison Selection in a Three-Choice "Equivalence" Test

			At least one "correct"	At least two "correct"	3 correct	Different choice for each sample
C1-A1	C2-A1	C3-A1	X			
C1-A1	C2-A1	C3-A2	X			
C1-A1	C2-A1	C3-A3	X	X		
C1-A1	C2-A2	C3-A1	X	X		
C1-A1	C2-A2	C3-A2	X	X		
C1-A1	C2-A2	C3-A3	X	X	X	X
C1-A1	C2-A3	C3-A1	X			
C1-A1	C2-A3	C3-A2	X			X
C1-A1	C2-A3	C3-A3	X	X		
C1-A2	C2-A1	C3-A1				
C1-A2	C2-A1	C3-A2				
C1-A2	C2-A1	C3-A3	X			X
C1-A2	C2-A2	C3-A1	X			
C1-A2	C2-A2	C3-A2	X			
C1-A2	C2-A2	C3-A3	X	X		
C1-A2	C2-A3	C3-A1				X
C1-A2	C2-A3	C3-A2				
C1-A2	C2-A3	C3-A3	X			
C1-A3	C2-A1	C3-A1				
C1-A3	C2-A1	C3-A2				X
C1-A3	C2-A1	C3-A3	X			
C1-A3	C2-A2	C3-A1	X			X
C1-A3	C2-A2	C3-A2	X			
C1-A3	C2-A2	C3-A3	X	X		
C1-A3	C2-A3	C3-A1				
C1-A3	C2-A3	C3-A2				
C1-A3	C2-A3	C3-A3	X			
	27		19	7	1	6